Single-cell transcriptomics identifies drivers of local inflammation in multiple sclerosis

David Schafflick[1,†], Michael Cole[2,†], Maike Hartlehnert[1,†], Tobias Lautwein[1], Konrad Buscher[4], Jolien Wolbert[1], Sven G. Meuth[1], Mark Stettner[5], Christoph Kleinschnitz[5], Tanja Kuhlmann[6], Catharina C. Gross[1], Heinz Wiendl[1], Nir Yosef [3,7,8,*,‡], and Gerd Meyer zu Horste[1,*,‡]

[1]Department of Neurology with Institute of Translational Neuroinflammation, University Hospital Münster, Münster, Germany.

[2]Department of Physics, University of California, Berkeley, CA, USA.

[3]Department of Electrical Engineering & Computer Science, Center for Computational Biology, University of California, Berkeley, CA, USA.

[4]Department of Nephrology, University Hospital Münster, Münster, Germany.

[5]Department of Neurology, University Hospital Essen, Essen, Germany.

[6]Department of Neuropathology, University Hospital Münster, Münster, Germany.

[7]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA.

[8]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

†*These authors contributed equally*

‡*These authors jointly supervised the study*

**Correspondence to:*

Gerd Meyer zu Hörste, MD
Department of Neurology
University Hospital Münster

Albert-Schweitzer-Campus 1, Bldg A1
48149 Münster
Germany
Tel.: +49 251 83 44428
Fax.: +49 251 980 2812
gerd.meyerzuhoerste@ukmuenster.de

Nir Yosef, PhD
Department of EECS
University of California, Berkeley

378 Stanley Hall
Berkeley 94720
USA
Tel.: +1 510 642 9640
Fax.: +1510 643 7846
niryosef@berkeley.edu

Running title: CSF cell transcriptomics in MS

Keywords: transcriptomics, single-cell RNA-seq, cerebrospinal fluid, multiple sclerosis, T follicular helper cells, experimental autoimmune encephalomyelitis.

43 **One Sentence Summary**

44 Unbiased single-cell transcriptomics re-defines the transcriptional landscape of cerebrospinal fluid
45 leukocytes and identifies T follicular helper cells as essential drivers of local inflammation in multiple
46 sclerosis.

47

48 **Abstract**

49

50 Single-cell transcriptomics enables unbiased biological discovery and holds new promise for
51 personalized medicine. However, its potential for understanding human diseases by comparing patient
52 vs. control samples in a clinical setting remains largely unexplored. Here, we applied single-cell RNA-
53 sequencing (scRNA-seq) to rare cerebrospinal fluid (CSF) specimens from well-characterized controls
54 and patients with multiple sclerosis (MS) – a prototypic inflammatory disease of the central nervous
55 system (CNS). We thereby generated and validated the first transcriptional atlas of single CSF
56 leukocytes in health and disease. In MS patients, we found an expansion of natural killer cells and late
57 B cell lineages and based on these insights we developed a score with potential diagnostic relevance.
58 Using this analytical approach, we identified and characterized activated phenotypes of MS-derived
59 CSF leukocytes, including an enrichment in T follicular helper (TFH) cell transcriptional signatures.
60 We validated the expansion of such B cell-helping TFH cells in MS patients and demonstrated that TFH
61 cells exacerbate symptoms in an animal model of MS and promote B cell infiltration of the CNS. TFH-
62 dependent B cell expansion may thus drive local CNS autoimmunity in MS. Our study demonstrates
63 how single-cell transcriptomics can identify novel disease mechanisms in a clinically-relevant case-
64 control study design.

65

Introduction

Single-cell transcriptomics is a transformative and rapidly evolving technology generating biological information at unprecedented resolution and scale. The technique has mostly been employed to re-define the heterogeneity of complex tissues derived from healthy rodents or humans *(1, 2)*. The novelty of these studies has mostly been limited to the identification of previously unrecognized cell types or cell phenotypes *(3)* and the regulation of their development. Diseased tissues have also been analyzed with single-cell technologies and the cancer field has seen especially rapid adaptation of these methods *(4, 5)*. Proponents of the technology posit that insights from single-cell transcriptomics are likely to translate into palpable benefits for human patients and enable precision medicine in the not-too-distant future *(6–8)*. However, outside of the field of cancer *(9)*, we are currently aware of only a handful of studies that utilize this technology to compare tissue samples from disease-affected donors against those of separate control donors in a clinically relevant setting *(10, 11)*. This leaves many methodological questions unexplored. Case-control studies are particularly important in systemic immune disease, when healthy control tissue cannot be reliably obtained. Indeed, many analytical tools for identifying differences between two sets of single-cell profiles (e.g. malignant vs. non-malignant) have been developed *(12)*, but their applicability to a clinically relevant case-control scenario has not yet been examined.

Here, we applied single-cell transcriptomics to cerebrospinal fluid (CSF) cells from patients with multiple sclerosis (MS) and controls, validating key findings with flow cytometry and mouse model studies. MS is a chronic inflammatory, demyelinating disorder of the central nervous system (CNS) – most likely of autoimmune origin – causing substantial disability *(13)*. We chose this paradigmatic inflammatory disease, because many questions remain unanswered despite a vast amount of available literature. Evidence supports the involvement of both T cells and B cells in MS, but the relative contribution of each cell type to disease aetiology is unknown. On the one hand, both the expansion of B cells and the production of immunoglobulins occur in the CNS *(14)* and B cell depleting therapies are effective in MS *(15)*. On the other hand, T cells are abundant in MS lesions *(16, 17)* and T cells are affected by many established MS treatments and induce an MS-like condition named *experimental*

autoimmune encephalomyelitis (EAE) in rodents *(18)*. Much needs to be learned about the interaction of T with B cells in MS.

CSF is a rare and clinically important specimen that has been studied extensively in MS, but has not yet been adequately analyzed with unbiased transcriptome methods *(19)*. We speculated that a study of this fluid can serve as the basis for an important proof of principle: translating single-cell transcriptomic to the bedside. CSF is a clear liquid that envelops the CNS and provides mechanical protection and trophic support *(20)* and acts as transport medium for immune cells *(21)*. Under healthy conditions, the *non-cellular* fraction of CSF is mostly an ultra-filtrate of serum *(22)*. In contrast, CSF *cells* - derived exclusively from the hematopoietic lineage - exhibit a distinct and tightly controlled cellular composition. Compared to blood, leukocyte concentrations in the CSF are 1,000-fold lower and $CD4^+$ T lymphocytes predominate, while myeloid-lineage cells are rare *(23)*. Clinically, CSF provides a unique diagnostic window into immune-related processes in the CNS. In MS, CSF exhibits several disease-associated changes including an increased concentration of oligoclonal immunoglobulins *(24, 25)*. Flow cytometry-based studies have also identified an expansion of B lineage cells in MS *(23, 26)* with evidence of antigen-driven maturation *(27, 28)*. However, the mechanisms promoting maturation of B cells such as class-switching in the CSF have not been identified and an unbiased characterization of CSF cells is missing.

Here we have used single-cell transcriptomics to generate a comprehensive map of the cellular composition and transcriptional phenotype of CSF cells in MS, demonstrating the feasibility of this technique in its application to human CSF samples. We find high levels of transcriptional and cellular heterogeneity across donors, an important consideration for future power calculations. We demonstrate why analyses aimed at capturing relevant disease-associated changes across a transcriptional continuum require novel analytical tools, and we introduced a new approach termed cell set enrichment analysis (CSEA) to address some of these challenges. Through these analyses, we make disease-related discoveries, like class-switched B lineage cells expanded in the CSF in MS. These changes coincide with an expansion of B cell-helping T follicular helper (TFH) cells that promote CNS auto-immunity and local B cell infiltration in animal models of the disease. These insights, derived from single-cell

transcriptome technology, lead us to propose a new cellular mechanism, locally driving CNS autoimmunity and disability in MS.

Results

Single-cell transcriptomics identifies the composition of cerebrospinal fluid cells

We aimed to characterize clinically relevant CSF cells in greater detail and to evaluate the applicability of single-cell transcriptomics in a translational setting. We optimized processing of primary human CSF cells (Methods) decaying rapidly in nutrient-poor CSF *(29)* and analysed these cells using 1) microfluidics-based single-cell RNA-sequencing (scRNA-seq) *(30)* and 2) flow cytometry (Fig. 1A). We first performed scRNA-seq on total unsorted CSF cells from treatment-naïve patients (n = 6) with either a first episode indicative of MS (i.e. clinically isolated syndrome (CIS)) or a first diagnosis of relapsing-remitting MS. For simplicity, we refer to this cohort as MS (Methods). Patients with idiopathic intracranial hypertension (IIH) served as controls (n = 6), because CSF itself is normal in IIH *(31)* while the production and absorption of CSF are unbalanced *(32)*. Both cohorts were well matched with regard to age and sex (Fig. S1A and Table S1). Standard CSF and disease parameters were either comparable between groups or exhibited known MS-associated changes (Fig. S1B and Table S2).

After quality control and removal of low quality cells and samples (2 donors per group; see Methods), our scRNA-seq approach returned transcriptional information for a total of 22,357 high-quality CSF cells from 4 control and 4 MS donors, with an average of 833 ± 193 SD genes detected per cell (Table S3). After normalization (Methods) and unbiased cell type clustering, we identified a total of 10 CSF cell clusters (Fig. 1B and Fig. S2A). Initially, $CD4^+$ T cells did not cluster reliably into known lineages and were therefore tentatively merged into one cluster (CD4_Tc). We manually assigned cluster identities based on known marker gene expression (Fig. 1C-D and Table S4; see Methods) and gene set enrichment analysis (GSEA) of marker genes (Table. S5). CSF cells featured a strong predominance of T cells (more $CD4^+$ than $CD8^+$) over monocyte lineage cells, natural killer (NK) cells, dendritic cells (DC), and B lineage cells including B cells (Bc*)* and plasma cells (plasma*)* (Fig. 1B and Fig. S2A-B). Granulocytes, megakaryocytes, and non-hematopoietic cells (e.g. neurons, glia, ependymal cells) were

not represented in our clustering of CSF cells (Fig. 1B) as these cell types are not present in the CSF *(23)*. Simultaneous flow cytometry of samples from all cohorts confirmed this unique composition of CSF leukocytes (Fig. S3A-C) in accordance with previous studies *(23)*. Thus, single-cell transcriptomics reliably reconstructs the composition of primary human CSF cells.

Single-cell transcriptomics identifies an MS-specific composition of CSF leukocytes

Next, we analysed our dataset for disease-specific differences in CSF cell composition (Fig. 2A). Overall, inter-donor variability was high (Fig. S2B). Despite this variability, there were significant compositional differences between the MS and control cohorts (Methods). Binomial regression modelling of scRNA-seq cluster membership counts reflects a significant decrease in the proportion of non-classical monocytes relative to classical monocytes in MS (Wald test $P < 10^{-8}$; Fig. 2B). A decreased ratio of non-classical / classical monocytes was confirmed by flow cytometry (t-test $P < 0.01$; Fig. S3B). The absolute abundance of non-classical monocytes is known to decrease in MS *(33)*; in our small scRNA-seq study, high variability in absolute non-classical monocyte abundance across the 8 donors suffices to explain apparent sampling differences between disease conditions ($P > 0.01$, empirical Bayes moderated t-test). Despite high inter-donor variability, we found statistically significant expansions of NK cells, B cells, and class-switched late lineage B cells (i.e. plasma cells) in MS ($P < 0.01$ empirical Bayes moderated t-test; Fig. 2A-C) that was confirmed by flow cytometry (t-test $P < 0.01$; Fig. S3B) and was in accordance with previous studies *(33, 34)*. Of note, plasma cells were detected in samples of all 4 MS patients but were virtually absent from control-derived CSF samples (Fig. 2A and Fig. S2B).

The expansion of B lineage cells was a uniquely MS-specific feature (Fig. S4C) and we therefore examined these clusters in greater detail. In the B cell cluster, IGHD (marker of naïve B cells) and IGHM genes were dominantly expressed in 5% and 34% of B cells, respectively (Fig. S4A,D). The expression of heavy chain genes was dominated by IGHG genes in the plasma cell cluster (83%; Fig. S4B,D) while fewer cells expressed IGHA genes (encoding IgA chains). This verifies that the vast majority of plasma cells in the CSF are class-switched. In both B cells and plasma cells the ratio between

dominant κ-light chain (encoded by IGKC) expression and dominant λ-light chain (encoded by IGLC genes) expression was approximately 2:1 – a physiological surface expression ratio for blood B cells (Fig. S4D). In accordance with previous studies *(14, 27, 28, 35)*, our findings suggest that local B cell maturation, including both class-switching and proliferation, occur within the CSF compartment in MS. Our comprehensive profiling of CSF cells in MS had identified changes in the relative abundances of B lineage cells, NK cells, and in monocyte subsets. We speculated that a combination of these parameters – rather than one single parameter – could aid diagnosing MS if quantifiable with flow cytometry. We therefore used the combined flow cytometry data as a baseline for calculating a composite score that was higher in MS-derived than control CSF samples (Fig. S5). This score discriminated MS from control with good sensitivity and specificity in this preliminary cohort. This indicates that single-cell transcriptomics of CSF cells can propose novel diagnostic schemes.

Characterizing the distribution of CD4$^+$ T cell states

We next aimed to dissect the composition of the tentatively merged CD4$^+$ T cells in our data. We extracted all cells assigned to the CD4$^+$ T cell (CD4_Tc) cluster, performed secondary normalization and clustering, and thereby identified eight sub-clusters (Fig. 2D,E). Two of these were identified as remaining CD8$^+$ T cells (r-CD8) and remaining monocytes (r-mono) based on transcriptional markers and were removed from further analysis. A transcriptionally distinct (Fig. 2F and Table S4) cluster of FOXP3 expressing (i.e. most likely regulatory) T cells (Treg) was more abundant in two and especially abundant in one, albeit not all MS donors (Fig. S2C). Clinical and MRI disease features were not different in these two MS patients (data not shown) supporting sub-clinical MS heterogeneity. Based on mean expression and statistically significant one vs. all differential expression of known marker genes (Fig. 2F and Table S4) and cluster specific GSEA (Table S5), two of the remaining clusters were transcriptionally best described as naïve (up-regulation of SELL (CD62L), CCR7 (CD197), and CD27, FDR < 0.05; n_CD4: SELL^{hi}CCR7^{hi}CD44^{lo}CD69^{lo} and CD27^{hi}) and as proliferating or differentiating CD4$^+$ T cells (up-regulation of SELL, CCR7, FDR < 0.05; prol_CD4: SELL^{hi}CCR7^{hi}CD44^{lo}CD69^{lo} and CD27^{lo}). The latter cluster expressed ribosomal genes (e.g. *RPS8*, *RPS6*) and nucleus forming

transcripts. Abundance of such proliferating T helper cells was increased in MS-derived samples (Fig. 2G) potentially indicating local expansion of $CD4^+$ T cells in the CSF in MS. Three of the remaining clusters exhibited a memory-like phenotype ($SELL^{int/lo}CCR7^{int/lo}$) and were transcriptionally best described as central memory (up-regulation of CD69, FDR < 0.05; cm_CD4: $CD69^{hi}CD44^{hi}$ and $CD27^{hi}$), as early effector memory (up-regulation of IL7R and CD69, FDR < 0.05; eem_CD4: $CD69^{hi}CD44^{hi}$ and $CD28^{hi}$), and as late effector memory (up-regulation of IL7R, FDR < 0.05; lem_CD4: $CD69^{int}CD44^{hi}$ and $CD28^{lo}$) $CD4^+$ T cells. These clusters showed no significant disease-specific expansion or contraction, after accounting for donor variability. Flow cytometry detected no significant differences in the proportion of total $CD4^+$ vs. $CD8^+$ T cells in MS (Fig. 2H) indicating that changes to T cells in MS are subtle, occurring at the subset level.

While the division of the $CD4^+$ T cells into sub-clusters was informative in this context, we found that the resulting clusters are not very well distinguished from one another (Table S4) and that $CD4^+$ T cells *transcriptionally* instead form a continuum of cell states, in accordance with previous scRNA-seq studies *(30, 36)*. Indeed, independent of our clustering analysis, we explored how transcriptional signatures vary across the entire $CD4^+$ T cell population, using VISION (an updated R version of FastProject *(37)* https://github.com/YosefLab/VISION). This analysis highlighted a continuum of transcriptional $CD4^+$ T cell states that span multiple sub-clusters in terms of T cell activation and memory (Fig. S6A,B), thereby providing a view of the data complementary to our analysis above. Analysis of MS-related transcriptional changes of $CD4^+$ T cells may therefore benefit from techniques that do not depend on data-driven partitions, e.g., clusters. These insights motivated our development of CSEA below (see also Fig. S7,8).

Single-cell transcriptomics can help interpreting MS genetics, transcriptomics and diagnosis

We aimed to systematically compare our transcriptome characterization of CSF cells against available data-sets. A single study had previously reported expression profiling of CSF cells in relapsing-remitting MS, albeit not at single-cell level *(19)*. This approach identified signs of local B cell expansion, but offered limited additional insight because *unsorted* cells were profiled. We therefore

used our scRNA-seq data to systematically infer the cellular composition of these *unsorted* CSF cells in MS patients in relapse and remission (n = 26 per group) using a deconvolution algorithm *(38)*. Deconvolution was unable to reliably discern NK cells and most $CD4^+$ T cell subsets (Table S10), most likely due to the high transcriptional similarity between subsets. However, it was able to infer an increased abundance of cells resembling plasma cells and Tregs as well as decrease of non-classical monocytes in the CSF of MS patients (Fig. 3A). These results therefore support some of our findings in this independent cohort of MS patients. Furthermore, this demonstrates that tissue-specific scRNA-seq can help interpret available bulk-level patient-derived data-sets.

The immune cell type(s) causing or promoting MS remain subject of debate. Results from genome-wide association studies have often been interpreted to reflect T cell-dependent mechanisms driving MS *(39)*. We therefore systematically evaluated ~170 known genes associated with genetic MS risk loci *(40)* against their respective expression levels in the CSF cell clusters we had identified (Fig. 1). We found that a minority (17%) of MS risk genes were expressed in multiple clusters (e.g. *CD58*, *CD28*, *TYK2*) (Fig. 3B). Most MS risk genes were instead preferentially expressed in one or two clusters. Such genes with a 'cluster-enriched' pattern were mainly expressed in B cell and plasma cell clusters (19% of genes, e.g. *CD40*, *CXCR5*, *BACH2*), in NK cells (11% of genes, e.g. *MAPK1*, *TCF7*, *JAK1*), in pDCs (9%, e.g. *IKZF1*, *IRF8*), and in monocyte and mDCs (22%, e.g. *CD86*, *IFNGR2*). Notably, in CSF cells only 3% of MS risk genes showed highest expression in the $CD4^+$ T cell cluster (e.g. *FOXP1*, *SOCS1*, *IL7R*) and 14% showed enrichment in $CD8^+$ T cells (e.g. *BATF*, *ETS1*, *IZKF3*) (Fig. 3B). Although highest expression cannot be equated with highest functional relevance, our data suggest that multiple immune cell lineages in the CSF can be affected by genetic MS risk. This argues for a multi-lineage immune etiology of MS – potentially through the interaction of B lineage cells with other cell types.

Identifying and interpreting cluster-specific transcriptional changes in MS

After exploring the overall changes in the cellular composition of the CSF in MS, we next focused on each cell cluster individually, testing genes for up-regulation (FDR < 0.05) in MS (Table S4). The complete $CD4^+$ T cell (CD4_Tc) cluster as well as both $CD8^+$ T cell clusters exhibited an increased expression of

MHC class I genes (i.e. *HLA-A, HLA-C, B2M*) and of *IL32* in MS patients indicating increased activation *(23)*. In accordance, GSEA *(41)* showed enrichment of pathways associated with protein synthesis (e.g. peptide chain elongation; $P < 0.01$) and thus cellular activation in $CD4^+$ T cells and naïve $CD8^+$ in MS (Table S6). The $CD8^+$ T cell clusters showed higher expression of genes associated with activation and cytotoxicity (*GZMK, GZMA, PRF1* encoding perforin 1) and GSEA identified antigen presentation pathways in activated $CD8^+$ T cells ($P < 0.01$, Table S6). Overall, this suggests higher activation and cytolytic capacity of cytotoxic CSF cells in MS. Both classical and non-classical monocytes featured higher expression of genes associated with antigen presentation (e.g. *CD74, HLA-DRB1*) and with migration (e.g. *ITGB2* encoding integrin-β2). The non-classical monocyte cluster also showed signs of increased secretory activity (e.g. induction of *GRN* encoding granulin) and GSEA found antigen presentation and interferon signaling pathways enriched in this cluster ($P < 0.01$, Table S6). The mDC cluster showed an increased expression of MHC class II (i.e. *HLA-DRA*) and MHC class I genes (e.g. *HLA-A*) and induction of CD1E. This indicates a propensity for lipid antigen presentation. GSEA identified lymphocyte costimulation pathways in this cluster ($P < 0.01$, Table S6). We did not observe statistically significant disease-specific transcriptional changes in the NK, pDC, and Bc clusters. This may – at least in part – be due to low cell numbers in these clusters. Because plasma cells were virtually undetected in control patients estimation of differential expression effect was prohibited. In conclusion, our analysis of transcriptional changes individually in each cell cluster reflects an ongoing immune cell activation in the CSF in MS.

Cell set enrichment analysis helps identifying disease-specific transcriptional changes

Our approach above used conventional single-cell analysis steps: 1) identifying cell clusters, 2) obtaining differentially expressed (DE) genes between disease-states for every cluster, 3) using GSEA to test for over-representation of known gene-sets and ascribe biological meaning. We speculated that this approach would be particularly insensitive to gene signatures or cell states that are poorly represented by tight clustering – as observed in $CD4^+$ T cell subsets. For example, a certain functional property may be specific to MS but only be present in a small subset of cells within a cluster; these patterns could be easily missed in a cluster-wide MS vs. control comparison. We therefore developed a

novel procedure – cell set enrichment analysis (CSEA) – which reutilizes the GSEA test for working on ranked lists of cells rather than genes (Methods, Fig. S7). In this procedure, cells in a cluster are first ordered by a transcriptional phenotype of interest (e.g., summed expression of genes in a pathway). The statistical test can then detect cases in which a subset of cells from one group (e.g., MS) exhibit unusually high or low values of that transcriptional phenotype compared to cells from the second group (e.g., control). We refer to these exceptional groups of cells as *core cell sets*.

We used this technique for a more comprehensive and clustering-free exploration of disease-specific transcriptional changes in the $CD4^+$ T cell compartment. As a source for transcriptional phenotype, we used signature scores from the VISION pipeline (Fig. S7). VISION signature scores are calculated by summing the expression of specific sets of genes, which can reflect a dichotomy between conditions of interest (e.g., naïve vs. memory T cell state) or a certain cellular function (e.g., signaling through interleukin (IL)-2; see Methods). The gene signatures were obtained from databases such as MSigDB and NetPath *(42, 43)* and are based on literature curation and on mining of large numbers of published microarray and RNA-seq studies (Methods).

Our CSEA testing procedure returned lists of core cell sets driving statistically significant signature enrichments in MS ($P < 0.01$, Bonferroni adjusted, Table S7). We identified *core MS cell sets* (Methods) driving enrichments for both an exhausted-versus-naïve CD4 signature *(44)* and a memory T cell signature *(45)* (Fig. S8), both exhibiting considerable overlap with the memory sub-clusters in Figure 2D. Importantly, the memory cell clusters exhibited no significant MS-specific differential abundance in our standard analysis above, but CSEA highlights a subset of these cells with pronounced memory or exhausted phenotypes that are particularly abundant in MS CSF. This argues for persistent T cell activation in the CSF in MS. We further identified several other MS core cell sets with exceptionally high expression of transcriptional signatures of T helper cell (Th)1 *(46)*, induced (i)Treg *(47)*, and T follicular helper (TFH) cells *(48, 49)*. Importantly, the cells in each of these three core sets do not significantly cluster in a transcriptome- wide analysis (VISION consistency testing *P-value* > 0.1), suggesting that cluster-based analyses are not well suited for capturing this layer of cell phenotype; e.g., cells expressing a Th1-polarized transcriptome are spread across both naive and memory clusters. Our

316 novel analytical approach can therefore decouple clustering of cells from disease-state enrichment of
317 cells, providing a new framework for interpreting complex scRNA-seq datasets.
318 Overall, these CSEA results emphasize an expansion of $CD4^+$ T cells with a Treg, Th1, and TFH
319 phenotype in MS. The Th1 result could indicate a greater role for Th1 versus Th17 in MS disease in the
320 CSF. Interestingly, TFH cells are known to drive B cell maturation. This lead us to hypothesize that an
321 increase in TFH abundance is responsible for the differences we observed in the B cell compartment of
322 the CSF.
323
324 *Expansion of B cell-helping T follicular helper cells in the CSF in MS patients*
325
326 Our unbiased approach had identified MS-specific changes in the CSF: 1) increased numbers of class-
327 switched B cells, 2) induction of transcriptional indicators of B cell maturation within B cell clusters,
328 and 3) enrichment of signatures of B cell-helping *(50)* TFH cells in CSEA. We therefore next tested
329 whether TFH cells are in fact altered in the CSF in MS. Increased numbers of circulating TFH cells had
330 previously been described in the blood of MS patients *(51, 52)*. We found the proportion of
331 $CD3^+CD4^+CXCR5^+$ TFH cells (Fig. 4A) significantly increased in the CSF of MS patients (Fig. 4B and
332 Table S4). The proportion of activated TFH cells expressing PD-1 and ICOS was also increased in MS
333 (Fig. 4B) while the alternative $CD4^+CXCR5^-PD\text{-}1^+$ subset *(53)* was unchanged (data not shown)
334 suggesting that these are *bona fide* TFH cells. The abundance of activated TFH cells positively
335 correlated with the proportion of CSF plasma cells (Fig. 4C) suggesting that both subsets may be
336 functionally related in the CSF.
337 Next, we characterized CSF-resident TFH cells in greater detail by performing bulk RNA-sequencing
338 (Methods) of TFH cells sorted from the CSF of new cohorts of MS patients and controls. MS-specific
339 transcriptional changes were comparably subtle and no individual genes reached gene-level significance
340 for differential expression (Table S8). This indicates that numerical differences in TFH cell abundance
341 are more pronounced than transcriptional changes of TFH cell phenotype. To investigate this further,
342 we performed GSEA and found an enrichment of gene sets associated with T helper cell memory and
343 pathogenicity in MS-derived TFH cells ($P < 0.01$, Bonferroni correction; Table S9). Genes often

recurring in these enriched gene sets (Fig. S9) were associated with cytotoxicity and cell death (e.g. *GZMA, GZMK, CASP3, CASP4*) and with co-inhibitory function (e.g. *KLRG1, TIGIT, CTLA4*). In accordance with our CSEA results, this suggests that pathogenic TFH cells expand in the CSF in MS patients. TFH cells are essential for the maturation of plasma cells and memory B cells. TFH expansion may thus contribute to the local interaction between T and B cells and thus potentially drive the disease.

TFH cells promote B cell accumulation in the CNS in an animal model of MS

As a test to this hypothesis, we next evaluated the *in vivo* functional relevance of TFH cells using a common animal model of MS. We generated mice with deficiency of Bcl6 – the lineage-defining transcription factor of TFH cells *(50)* – restricted specifically to T cells. Such $CD4^{Cre}Bcl6^{fl/fl}$ mice lack TFH cells and fail to mount antigen-specific B cell responses, while differentiation of other T helper cell lineages is unaffected *(54)* (Fig. S10). The course of EAE – an animal model of MS – has not been investigated in these mice before. We therefore induced EAE using myelin oligodendrocyte glycoprotein $(MOG)_{35-55}$ peptide. EAE severity was significantly reduced in $CD4^{Cre}Bcl6^{fl/fl}$ mice compared to Cre-negative littermates (Fig. 5A). Accordingly, the number of inflammatory lesions and infiltrated area in the spinal cord of $CD4^{Cre}Bcl6^{fl/fl}$ mice was lower than in controls (Fig. 5B,C). When we extracted leukocytes infiltrating the CNS at the peak of EAE we found that the proportion of pro-inflammatory IL-17 producing $CD4^+$ T cells was reduced in $CD4^{Cre}Bcl6^{fl/fl}$ mice indicating a lower degree of CNS tissue destruction and inflammation in the absence of TFH cells (Fig. 5D). Next, we tested how the absence of TFH cells influenced B cells in the CNS and found a lower proportion of total B cells ($B220^+CD3^-$) infiltrating the CNS in $CD4^{Cre}Bcl6^{fl/fl}$ mice by flow cytometry (Fig. 5E). We also histologically stained for B cells in the CNS and again found a lower number of intraparenchymal B cells in the spinal cord of $CD4^{Cre}Bcl6^{fl/fl}$ mice with EAE compared to Cre-negative littermates (Fig. 5F). Taken together our data indicate that TFH cells enhance MS-like autoimmunity by locally supporting the expansion of B cells in the CNS.

Discussion

In this study we applied single-cell transcriptomics to rare and clinically relevant CSF specimen from MS patients and controls. We thereby create the first comprehensive map of the cellular composition and transcriptional phenotype of CSF cells. In analysing our data, we observed that – transcriptionally – $CD4^+$ T cells are best described as a continuum of cell states rather than clearly defined subsets or clusters *(36)*. This observation together with considerable inter-donor heterogeneity necessitated development of CSEA that facilitates extracting disease-specific mechanisms from complex scRNA-seq data, by focusing (in a data- driven way) on the most relevant subsets of cells. Beyond the specific application in this paper, we therefore expect that methods such as CSEA will be essential for conducting future single-cell transcriptomics studies with a case vs. control design.

Pooling CSEA together with other components of our analysis, we identified a potential multi-lineage immune aetiology of MS with an expansion of matured B lineage cells, NK cells, and proliferating T helper cell subsets in the CSF. Transcriptionally, MS-derived cells featured an enrichment of Th1- and TFH-like signatures, which formed the basis for speculating that TFH cells play a role in MS. In fact, we found that TFH cells accumulate in the CSF in MS and correlate with plasma cell numbers in accordance with a previous study *(55)*. We also found that TFH cells promote disease severity and local B cell expansion in an animal model of MS. In conjunction with a previous study *(56)*, our data provide strong *in vivo* evidence that a pathological interaction between TFH cells and B cells drives CNS inflammation. This sequential approach exemplifies how single-cell transcriptomics can be translated to the bedside and reverse validated in corresponding animal models. By translating technology, we here identify a new cellular mechanism, locally driving CNS autoimmunity and disability in MS.

Previous studies have shown that B cell clones at least partially expand in the CSF in MS *(27, 57)* while migration from the periphery also occurs *(14, 35)*. An importance of B cells in MS was previously suggested by the presence of oligoclonal immunoglobulins, by the expansion of plasmablasts in the CSF *(25, 58)*, and by the efficacy of B cell-depleting therapies in MS *(15)*. However, strong *in vivo* data

confirming a functional link between TFH cells and B cells in neuro-inflammation was not previously reported. Notably, the gene encoding the TFH marker CXCR5 is a genetic risk locus for MS *(59)*. Previous studies suggest that TFH cells and B cells in the CSF could be derived from meningeal sources. Chronic ongoing CNS inflammation induces ectopic lymphoid tissue (eLT) in the affected tissue in many autoimmune diseases and is thought to be the site of local auto-antibody production *(60)*. In MS, eLT develops in the meninges *(61, 62)*, contains B cells and TFH cells *(56, 63)*, and is located in close vicinity of degenerating axons and neurons *(60, 64–66)*. It remains to be tested experimentally, whether CSF and meningeal immune cells communicate and interact.

Finding controls for CSF-based studies is difficult. The ideal controls would be healthy and matched for all confounders *(67)*. However, lumbar punctures (LP) cannot be performed for solely scientific purposes in healthy volunteers. In addition, volunteers in clinical studies are usually males *(68)* while MS patients are predominantly female *(13)*. We therefore intentionally used IIH controls, which are well matched for sex, age and comorbidities (Fig. S1A) and CSF from IIH patients was found normal in a previous study *(26)*. Even in MS patients an LP is usually performed only once to exclude relevant differential diagnoses during the diagnostic work-up for a first relapse indicative of MS. We specifically recruited these *untreated first-relapse* patients for our study. Although this was not part of our formal inclusion criteria, we thereby enriched for patients currently in (first) relapse. The phenotype of CSF cells in remission may be different. We also intentionally limited our study to treatment-naïve patients since many MS treatments considerably impact peripheral or CSF leukocyte composition *(69)* or substantially alter the transcriptional profile of immune cells *(70)*. Characterizing treatment effects was not the focus of the present study. In fact, our study forms a reference point for future CSF transcriptomics studies in MS patients in other disease stages (e.g. remission or progressive) or while receiving disease modifying treatments.

Transcriptional studies in MS were initially performed in unsorted peripheral blood mononuclear cells, because these are easily accessible *(71)*. Some studies focussed on defined cell populations like T cells *(72)*, on gender-specific differences *(73)*, or correlated transcriptional findings with genetic information

(74). More recent and larger-scale studies also included different MS treatments *(75–77)*, or enriched for myelin antigen-specific T cells from the blood of MS patients *(78)*. Although these previous studies have provided important insights into peripheral immune responses in MS, they all feature essential inherent short-comings: 1) peripheral blood cells constitute a poor surrogate of inflammation in the brain in MS, 2) transcriptional studies using mixed populations cannot distinguish changes in cell composition from changes in gene expression per cell, 3) previous enrichment techniques solely focussed on T cells *(78)* – a hypothesis driven approach.

Our study provides the first unbiased and single-cell-resolution look at local immune processes in the CSF. On a wider perspective, our study demonstrates that scRNA-seq of human CSF cells can generate novel hypotheses about debilitating neurological diseases that can be validated using reverse-translational tools. Our study thus forms the basis for a future application of the method in other neurological diseases such as Parkinson's and Alzheimer's disease.

Materials and Methods

Patient recruiting and inclusion

A total of 26 treatment-naive patients with MS or clinically isolated syndrome (CIS) receiving a lumbar puncture (LP) for diagnostic purposes, were prospectively recruited (Table S1). The control group consisted of 22 patients diagnosed with idiopathic intracranial hypertension (IIH) (Table S1). Patients were recruited in three consecutive cohorts. CSF cells from cohort 1 were used for unsorted single-cell RNA-seq (6 IIH vs. 6 MS patients). CSF cells from cohort 2 were analysed by flow cytometry only (7 IIH vs. 11 MS patients), and cells from cohort 3 were flow sorted for RNA-seq of $CD3^+CD4^+CXCR5^+$ TFH cells (9 IIH vs. 9 MS patients) (Table S1 and Fig. S1). All patients were of Caucasian ethnicity and gave written informed consent. The study was performed in accordance with the declaration of Helsinki and approved by the local ethics committees under reference number 2015-522-f-S.

For MS patients, formal inclusion criteria were defined as: 1) treatment naive patients with a first episode suggestive of MS (i.e. clinically isolated syndrome (CIS)) or with relapsing-remitting (RR)MS diagnosed based on MAGNIMS criteria *(79, 80)*, 2) patients receiving LP for diagnostic purposes and consenting to participate. Exclusion criteria for MS patients were defined as: 1) questionable diagnosis of MS by clinical signs or magnetic resonance imaging (MRI) findings, 2) secondary chronic progressive MS or primary progressive MS. IIH patients were included, if they gave informed consent. Exclusion criteria for all patients were: 1) immunologically relevant co-morbidities (e.g. rheumatologic diseases), 2) severe concomitant infectious diseases (e.g. HIV, meningitis, encephalitis), 3) pregnancy or breastfeeding, 4) younger than 18 years, 5) mental illness impairing the ability to give informed consent, 6) artificial blood contamination during the lumbar puncture resulting in >200 red blood cells / µl. patients whose diagnostic work-up revealed a diagnosis other than MS / IIH within four weeks of clinical follow-up were retrospectively excluded (Fig. S1C). The following diagnostic tests were performed in all MS patients to exclude differential diagnoses: PCR for cytomegaly virus, Ebstein-Barr virus, Human Herpes Virus-6, Herpes simplex Virus (HSV)-1, HSV-2 and Varicella-Zoster Virus in CSF. Blood tests for anti-HAV IgM, HBsAg, anti-HBc, anti-HCV, rheuma factor, Waaler-Rose Test, anti-cyclic citrullinated peptide (CCP), antinuclear antibody (ANA), anti-double strand (ds)DNA

antibodies, antineutrophil cytoplasmic antibodies (ANCA). CSF and serum were tested by the Treponema pallidum hemagglutination assay (TPHA). Borrelia burgdorferi was detected in CSF and blood by ELISA. R version 3.4.4 and RStudio 1.1.447 were used for the analysis of clinical and human flow cytometry data.

Sampling and flow cytometry analysis of cerebrospinal fluid cells

LPs were performed under sterile conditions using 20G Sprotte Canulae (Pajunk Medical). Up to 5 ml of CSF and 3 ml of blood were collected for scientific purposes in addition to diagnostic material. All samples were pseudonymised at collection. CSF was transported to further processing as quickly as possible and centrifuged at 300g for 10 min. The supernatant was removed and CSF cells were resuspended in 5 ml of X-Vivo15 media (Lonza) and stored at 4°C until further processing. CSF flow cytometry was performed in all donors using a Navious flow cytometer (Beckman Coulter). Cells were incubated in VersaLyse buffer and stained using the following anti-human antibodies (Beckman Coulter; clone names indicated): CD3 (UCHT1); CD4 (13B8.2); CD8 (B9.11); CD14 (RMO52); CD16 (3G8); CD19 (J3-119); CD45 (J.33); CD56 (C218); CD138 (B-A38).

For scRNA-seq, CSF cells in media were centrifuged at 400 g for 5 min and resuspended in 40 µl of X-Vivo15 media. 5 µl of the single-cell suspension was manually counted in a Fuchs-Rosenthal chamber. The maximum of CSF cells used for input was 10,000 cells. If total available CSF cell numbers were lower than 10,000 cells, all available cells were processed. On average 5,917 cells ± 1,505 SD (control 6,167 cells ± 2,614 SD vs. MS 5,667 cells ± 1,506 SD) CSF cells were used as input per donor.

A summed composite score differentiating flow cytometry results of MS form control patients was calculated. First, for each sample analysed by flow cytometry four ratios were calculated: 1) proportion of NK to $CD4^+$ T cells, 2) Bc to $CD4^+$ T cells, 3) $CD8^+$ to $CD4^+$ T cells, and 4) $CD14^+CD16^-$ to $CD14^+CD16^+$ monocytes. Each resulting group average was normalized to a value of 1 by dividing individual values by the group average. These four normalized ratios were added to obtain a basic composite score. A value of 1 was added to the basic composite in each case that an elevated immunoglobulin index or oligoclonal bands were detected in the samples; i.e. a maximum of 2 was added to the basic score. This extended composite score was named ‘Münster MS composite’ score.

Receiver operator curve (ROC) analysis of the composites was performed and the area under the curve (AUC) was calculated using the Glm and rocplot functions of the Deducer package v0.7-9 in R.

Generation of single-cell libraries and sequencing

Single-cell suspensions were loaded onto the Chromium Single Cell Controller using the Chromium Single Cell 3' Library & Gel Bead Kit v2 (both from 10X Genomics) chemistry following the manufacturer's instructions. Sample processing and library preparation was performed according to manufacturer instructions using AMPure beads (Beckman Coulter). Sequencing was carried out on a local Illumina Nextseq 500 using the High-Out 75 cycle kit with a 26-8-0-57 read setup. Average sequencing depth was 51,064 ± 13,041 SEM reads/cell (Table S3).

Preprocessing of sequencing data

The analysis pipeline for scRNA-seq data is illustrated in Fig. S11. Processing of sequencing data was performed with the *cellranger* pipeline v2.0.2 (10X Genomics) according to the manufacturer's instructions. Raw bcl files were de-multiplexed using the *cellranger mkfastq* pipeline. Subsequent read alignments and transcript counting was done individually for each sample using the *cellranger count* pipeline with standard parameters. The *cellranger aggr* pipeline was employed, to ensure that all samples had the same number of confidently mapped reads per cell. The *cellranger* computations were carried out at the High Performance Computing Facility of the Westfälische Wilhems-University (WWU) Münster. The pre-quality control (QC) total number of cells with available scRNA-seq profiles was 22,357 with an average of 3,176.3 ± 3,246.2 SD individual cells available per control donor and 2,413.0 ± 1,198.7 SD individual cells available per MS donor (Table S3). This corresponds to an average cell recovery rate of 48.8% ± 27.5% SD compared to input cells (control 46.1% ± 33.1% vs MS 51.6% ± 32.0%)

Single-Cell Sample Filtering

Initial exploratory data analysis identified one MS sample and one IIH sample whose clustering did not overlap with other samples (data not shown). This tight clustering suggested either strong batch effects

or significant contamination. Both samples were excluded from further analysis, leaving 5 control- and 5 MS-derived samples.

Nine barcode-level quality control (QC) metrics were computed for the unfiltered 10x Cell Ranger output: (1) number of unique molecular identifiers (UMIs), (2) number of reads, (3) mean reads per UMI, (4) standard deviation of reads per UMI, (5) percent of reads confidently mapped to the gene, (6) percent of reads mapped to the genome but not a gene, (7) percent of reads unmapped, (8) percent of UMIs corrected by the Cell Ranger pipeline, and (9) the number of cell barcodes corrected by the Cell Ranger pipeline. These metrics were used for filtering and normalization. We applied the gene and sample filtering using a scheme previously described *(81)*. This involved four steps:

1. Define *common genes* based on UMI counts: Genes with n_u or more UMIs in at least 25% of barcodes, where n_u is the upper-quartile of the non-zero elements of the UMI matrix.
2. Filter samples based on QC metrics. Remove samples with low numbers of reads, low proportions of mapped reads, or low numbers of detected common genes. The threshold for each measure is defined data-adaptively: A sample may fail any criterion if the associated metric under-performs by z_{cut} standard deviations from the mean metric value or by z_{cut} median absolute deviations from the median metric value. Here we have used $z_{cut} = 2$. This function is implemented in scone::metric_sample_filter (see below).
3. Remove barcodes from donors with fewer than 100 barcodes following sample filtering. These donors have contributed too few high-quality samples to reliably estimate donor-specific effects. Only seven cells were removed in this step.
4. Filter genes based on UMI counts: Genes with n_u or more UMIs in at least n_s barcodes, where n_u is the upper-quartile of the non-zero elements of the sample-filtered UMI matrix. We have set $n_s = 5$ to accommodate markers of rare populations. This sub-step ensures that included genes are detected in a sufficient number of samples after sample filtering. For the $CD4^+$-only analysis this step was applied again after the data matrix was subset to include only $CD4^+$ clusters.

Single-Cell Normalization

We utilized the SCONE package *(81)* to select an appropriate normalization based on a standardized panel of performance criteria. Clustering and Correlation evaluations were based on principal component analysis (PCA)-based dimensionality reduction to ten principal components.

Scaling normalization: We included a number of scaling methods with wrappers implemented in the SCONE package, including: no normalization, total count normalization, trimmed mean of M-values normalization method (TMM) normalization, upper quartile normalization, full quantile normalization, and the relative log expression normalization.

Categorical covariates: We considered normalization procedures that include a linear regression-based batch adjustment for log-transformed expression data. For our purposes we considered the donor ID as a batch covariate. Normalized matrices were scored for batch mixing using the SCONE batch silhouette score. We also monitored the silhouette score of MS vs. control status, although we never explicitly included this categorical biological covariate as part of the adjustment model. The stratified Partitioning Around Medoids (PAM) argument was applied to the evaluation of *de novo* PAM clusters, considering a range of *K* from 2 to 8.

Control genes: Positive controls were selected from the top 500 most common gene symbols referenced in the Molecular Signatures Database (MSigDB) C7 collection of immunological signatures *(42)*. Negative controls were selected from a previous study *(82)*. In order to match sets for mean expression, genes were binned according to the rounded mean $\log_2$ expression (adding 1 to each observation). Genes for the positive control set, and two negative control gene sets (adjustment and evaluation) were drawn in equal numbers (maximum) from each expression bin, for a total of 207 genes each. For the $CD4^+$-only analysis the lists were slightly smaller at 196 genes each.

Unwanted variation: We performed adjustment based on principal components (PC) of the QC (named qPC). Such qPC-based adjustment involved regression on PCs of the QC metrics discussed above. We also performed the remove unwanted variation (RUV) normalization strategy *(83)*. Both RUVg and qPC adjustments considered by SCONE were performed over a range of 0 to 8 factors.

Selected normalizations: The top performing normalizations for both the full analysis and $CD4^+$-only analysis both involve relative log expression scaling, qPC-based adjustment, and batch adjustment. For

581 the full analysis this normalization included all eight qPCs, whereas the T cell analysis included only
582 four.
583
584 *Seurat Analysis*
585 After sample filtering, we loaded the normalized log-transformed UMI matrix into the Seurat analysis
586 pipeline *(84)*. Following data scaling and PCA, we clustered the cells in the first ten principal
587 components using the *Seurat::FindClusters* function. Clustering resolution was set to 0.6. Identical
588 options were used for the $CD4^+$-only Seurat analysis (see below), defining subclusters of those cells.
589 We manually annotated clusters based on marker gene expression and enrichment analyses described
590 below. t-distributed stochastic neighbour embedding (t-SNE) data representations were computed using
591 the fast option in *Seurat::RunTSNE*.
592
593 *VISION Analysis*
594 We passed raw and normalized UMI data to the VISION pipeline
595 (https://github.com/YosefLab/VISION) *(37)*. Mean expression per gene symbol was calculated prior to
596 the analysis in order to make the features relatable to general gene signatures. The goal of FastProject
597 analysis – on which VISION is based – is to uncover biologically meaningful gene signatures that vary
598 coherently across single-cell neighbourhoods *(37)*. These signatures can help assign meaning to the
599 dominant expression differences between clusters. In addition to raw data, we passed QC, donor, status,
600 and Seurat cluster covariates for exploratory analysis and visualization. VISION quantifies the extent
601 to which cell signature values cluster across the cell manifold by using "consistency testing." VISION
602 scores the extent to which neighbouring cells (similar expression profiled) are predictive of a cell's
603 signature value using autocorrelation (Giri's C) statistics, comparing against random permutations in
604 order to assign statistical significance with respect to a uniform null model. We also included the Seurat
605 t-SNE as a precomputed projection. Our signature set includes:
606 1. Human cell cycle genes described before *(2)*, representing sets of genes marking G1/S, S,
607 G2/M, M, and M/G1 phases.
608 2. The MSigDB C7 immunological signature collection *(42)*.

3. T_H signatures compiled previously *(46)*.
4. NetPath database signatures *(43)*.
5. Curated T cell signatures *(36)*.
6. Curated T_{FH} *(48, 49)* signature sets.

Housekeeping genes were referenced from the same source as the SCONE negative controls above *(82)*.

Comparing gene expression and cluster composition between MS patients versus controls

Differential Composition Analysis

For both the initial and the $CD4^+$-only clustering, we used *limma::voom (85)* to test the difference in cluster abundances (cell counts) between MS donors and control donors.

Binomial regression modelling was applied to compare the relative sampling of classical and non-classical monocytes in monocyte fraction of MS and control CSF samples. The classical fraction of monocytes increased significantly from 17% in control donors to 32% in MS donors (Wald test $P < 10^{-8}$).

Cluster-specific expression analysis

We performed one versus all comparisons following each clustering analysis in order to annotate the clusters. One versus all differential expression (DE) tests *P*-values were used to rank genes by the extent they are up-regulated in one cluster over all others. Tests were performed separately for each donor sample with at least 10 cells in the target cluster. qPC factors used for normalization above were incorporated into a linear predictor for limma-voom DE testing. Results for each donor sample were combined in multiple ways, calculating median log fold changes, meta-analysis *P*-values for one-sided tests using Stouffer's method, and irreproducible discovery rates (IDR) *(86)* for two-sided tests using the est.IDRm tool in the scRAD package *(87)* for all genes and comparisons (Table S4). Examples of reproducible marker RNAs (FDR < 0.05, IDR < 0.05, and median FC > 2-fold) for the initial clustering can be found below:

Cluster (ID)	**Marker RNA Gene Symbol**
B cell (Bc)	CD79A, MS4A1
Plasma cell (plasma)	CD79A, XBP1
$CD4^+$ T cell (CD4_Tc)	IL7R
Naïve $CD8^+$ T cell (nCD8_Tc)	NKG7, CCL5
Activated $CD8^+$ T cell (aCD8_Tc)	NKG7, CCL5, GZMK
Natural killer cell (NK)	NKG7, GNLY
Plasmacytoid dendritic cell (pDC)	CLEC4C
Myeloid dendritic cell (mDC)	LYZ, FCER1A, CD1C
Classical monocyte (class_mono)	LYZ, CD14, S100A9
Non-classical monocyte (nc_mono)	LYZ, CD14, FCGR3A, MS4A7

After re-clustering the $CD4^+$ T cell cluster (CD4_Tc), the marker criteria above identified contaminating populations with markers of non $CD4^+$ T cell lineages. These cells were erroneously clustered together with $CD4^+$ T cells in the initial clustering and we named them remaining $CD8^+$ T cells (r-CD8) and remaining monocytes (r-mono).

Tc Contaminant Cluster (ID)	**Marker RNA Gene Symbol**
Remaining $CD8^+$ T cell (r-CD8)	NKG7, CCL5
Remaining monocytes (r-mono)	LYZ, FCER1A, CD1C, CD14, FCGR3A, MS4A7

Remaining $CD4^+$ subclusters were annotated by joint considerations of i) significant (FDR < 0.05) and large log2 fold change greater than 0.1 and ii) Mean expression of known markers (Fig. 2F).

Per cluster case-control comparison

Cluster-specific gene expression differences between MS and control were also assessed. Donors were only included in a comparison if 10 or more cells from the target cluster were detected in the donor's

sample. All pairings of MS donors with control donors were considered (up to 16). For each valid case-control pair, DE analysis was performed using limma-voom, as in the marker analysis, but comparing case cells against control cells. Log fold change was summarized by the median of log fold changes estimated across the donor pairs. Meta-analysis was performed on all possible pairings of cases and controls (up to 4! = 24); the median meta-analysis P-value was reported. IDR modelling was applied at the pair level, modelling the reproducibility of up to 16 replicate significance signals (Table S4). Some genes are very lowly expressed across individual clusters, resulting in unstable statistical estimation for those genes. Genes were filtered before DE if they had mean un-normalized UMI counts below 0.05.

Gene Set Enrichment Analysis (GSEA)

After deriving lists of differentially expressed genes, we sought to uncover enrichment for particular gene sets to capture biological differences between samples. We applied GSEA tests *(41)* to all single-cell differential expression tests returning cluster specific gene expression (i.e. genes expressed by one cluster vs. other clusters, Table S5) and disease specific gene expression (i.e. genes expressed within one cluster in MS cells vs. control cells, Table S6). We used signed significance scores based on meta-analysis *P*-values as gene signals and applied the Bonferroni adjustment to control the FWER for each category of hypotheses (i.e. test type, cluster, and sign). Sets considered in this analysis include all MSigDB C7 signature sets and all curated T cell signature sets described previously *(36)* with 10 or more genes quantified in the present study; “UP” and “DN” signature subsets were tested separately. The initial description of GSEA recommend simulating a null distribution for the GSEA test statistic at the gene level (e.g. recomputing lfcs for shuffled sample labels) *(41)*. This approach is computationally costly in our case; in this analysis we generated null distributions of the GSEA test statistic by shuffling gene set memberships, assigning empirical one-sided *P*-values based on simulation (https://CRAN.R-project.org/package=gsEasy).

Cell Set Enrichment Analysis (CSEA)

For the $CD4^{+}$-only analysis we considered a novel adaptation of the GSEA method, applying the technique to cell sets: CSEA (illustrated in Fig. S7). CSEA is a hypothesis testing method for

simultaneously uncovering enrichments and identifying subsets of cell sets of importance. In this procedure, a collection of cells is first ordered by a transcriptional phenotype of interest (e.g., sum expression of genes in a pathway). The resulting statistical test is sensitive to cases in which only a subset of cells from one group (e.g., MS) exhibit unusually high or low values of the transcriptional phenotype. The input to this method is a list of *N* cells, rank-ordered by some input signal. Our analysis uses VISION signature scores, reflecting known axes of biological variation. VISION signature scores – based on FastProject signature scores *(37)* – are computed by first centering and scaling each normalized log expression cell profile. Following scaling, the sum of gene expression values in the negative signature subset are subtracted from the sum of gene expression values in the positive signature subset. Signatures are normalized to the total number of genes in the set. For example, a signature set that describes a dichotomy between naïve and memory T cells may be used to score individual cells, indicating that some cells have higher expression of genes characterizing the naïve state and lower expression of genes characterizing the memory state. Using the notation previously described *(41)* we will use r_j to denote the cell j's signature score; indices have been sorted so that $r_j > r_{j+1}$ (alternatively in increasing order: $r_j < r_{j+1}$). The test involves considering all cells up to a specific position, i. A "hit" score is defined as the cumulative sum of signature score magnitudes (optionally exponentiated by parameter *p*: $|r_j|^p$) for members of cell set *S*, divided by the sum over all set members in the list. A "miss" score is similarly calculated for non-members of *S*, but without weighing by signature score magnitudes. The CSEA enrichment score (ES) is defined as the maximum of the difference between the running hit score and running miss score with respect to index i. When p=0, the ES reduces to a one-sided KS test statistic for differential signature analysis between cell sets. We apply the same permutation scheme as described for GSEA above. For p>0, CSEA cannot be seen as a simple differential signature test: CSEA tests for enrichment of a cell set at the high tail (or low tail) of the signature score distribution, but additionally weighs the set elements according to their signature value. This reduces the effects of low-magnitude cells in *S*, whereas all cells not in *S* are treated the same no matter the magnitude of their signature score. CSEA tests if high magnitude (positive or negative) cells are enriched at a specific tail, applying permutation tests to account for the additional variability induced by the magnitude weights. The set of indices up to where the objective score reaches its maximum also holds significance – in

GSEA *(41)* referred to as the "leading-edge" of the enrichment test. The intersection of the set *S* and the leading-edge is the *leading-edge subset*, representing an important core subset of cells driving an enrichment. For each VISION signature, we treated the computed signature scores as cell signature scores r_j. The sets under consideration were the mutually exclusive sets of MS and control cells. The goal of this approach is to identify core sets of cells that drive each biological condition's enrichment for high or low signature values (Fig. S7). Contaminating sub-populations in the CD4_Tc cluster, were removed prior to CSEA.

Bulk RNA-Seq of sorted TFH cells

CSF TFH cells were sorted on a BD FACS Aria™ III cell sorter using FACS Diva™ software following manufacturer's instructions using an 85 µm nozzle and the drop delay was determined using BD Accudrop™ beads. Sorting was performed using sort precision mode "purity" for live $CD3^+CD4^+CXCR5^+$ cells. Antibodies against PD-1 (EH12.2H7) and ICOS (C398.4A) were from Biolegend. Cells were sorted directly into 1,5 ml reaction tubes containing 100 µl RNA Lysis Buffer (Zymo Research). After sorting, tubes were vortexed, briefly centrifuged and frozen at -80 °C until RNA isolation. Data were analyzed using FlowJo software v10.4.1 (Tree Star, Inc.). Samples for bulk RNA-sequencing were prepared using a modified version of the SmartSeq2 protocol *(88)*. Unquantified purified RNA was used as input. Reaction volumes were scaled up and the number of PCR cycles during cDNA amplification adjusted accounting for the higher number of input cells compared to the original protocol *(88)*. Library Preparation was done by the Next UltraII FS DNA Library Prep Kit (New England Biolabs) using 1-3 ng of cDNA as input. Sequencing for 9 MS samples and 9 IIH samples was carried out on a NextSeq500 using the High-Out 75 cycle kit (Illumina).

Bulk expression quantification

RNA-seq reads were aligned to the RefSeq hg38 transcriptome (GRCh38.2) using Bowtie2 *(89)*. The resulting transcriptome alignments were processed using the RNA-Seq by Expectation Maximization (RSEM) toolkit to estimate expected counts over RefSeq transcripts *(90)*. Several genes were quantified multiple times due to alternative isoforms unrelated by RefSeq annotation. Before expression data

732 normalization, the gene entry with maximum counts was selected to represent the gene in further
733 analysis.
734
735 *Bulk data filtering*
736 Sample and gene filtering were similar to the scRNA-seq filtering method above, enforcing (> 107k
737 reads, > 10% read alignment (forced), > 93.3% common genes detected; corresponding to $z_{cut} = 20$). A
738 total of 5 samples were removed, leaving 13 samples. Setting $n_s = 1$, we analysed 11,383 genes below.
739 For each sample, we computed transcriptome alignment and quality metrics using FastQC (Babraham
740 Bioinformatics), Picard tools (Broad Institute), and custom scripts. Computed metrics included: (1)
741 number of reads; (2) number of aligned reads; (3) percentage of aligned reads; (4) number of duplicate
742 reads; (5) primer sequence contamination; (6) average insert size; (7) variance of insert size; (8)
743 sequence complexity; (9) percentage of unique reads; (10) ribosomal read fraction; (11) coding read
744 fraction; (12) UTR read fraction; (13) intronic read fraction; (14) intergenic read fraction; (15) mRNA
745 read fraction; (16) median coefficient of variation of coverage; (17) mean 5' coverage bias; (18) mean
746 3' coverage bias; and (19) mean 5' to 3' coverage bias.
747
748 *Bulk data normalization, unsupervised, and supervised analysis*
749 Data were normalized using SCONE. 569 positive controls were derived from MSigDB C7 entries
750 annotated to include TFH cell types, including the most frequently included gene symbols in those
751 entries. Negative controls for RUVg and evaluation were derived from the housekeeping gene list.
752 Control lists were sampled down to 186 genes per list so as to match mean expression of genes in each
753 list. The study group included two batches with 4/3 and 3/3 MS/IIH samples respectively. Biological
754 condition was used only for evaluation. SCONE recommended TMM scaling and adjustment for 2
755 factors of RUVg and batch condition.
756 We performed PCA on the scaled log-transformed normalized data for visualization. DE between MMS
757 and IIH donors was performed with limma-voom, using RUVg factors and batch in the model to adjust
758 for unwanted variation. Per-gene DE significance scores were computed from log-transformed *P*-values
759 and used for GSEA enrichment testing. Sets considered for testing included numbers 3,5, and 6

described in the VISION section. The 42 most frequent core members of the significant enrichments (Bonferroni adjusted *P*-value less than 0.01) – genes driving 7 or more of these enrichments – were selected and their normalized log values were correlated against each-other and represented in a sorted heatmap using *pheatmap* defaults.

Expression deconvolution using scRNA-seq data

Raw UMI mean counts per cluster were used as input for deconvolution. Cibersort was used for RNA expression deconvolution *(38)* on the E-MTAB69 dataset described previously *(19)*. We found that when using highly similar cell clusters as input for deconvolution (e.g., CD4_Tc together with $CD4^+$ T cell sub-clusters) lower abundance clusters (e.g., $CD4^+$ T cell sub-clusters) were not identified due to high transcriptional overlap. We therefore excluded the CD4_Tc cluster from deconvolution. A customized RNA signature was extracted based on the scRNA seq data (no quantile normalization, permutations 100, *Q*-value 0.1). UMI were transformed for correlation with microarray expression (x=log2(y+2)*1.5). Only correlations with $p < 0.05$ were used. The resulting signature contained 91 genes. A deconvolution of the original scRNAseq data served as control, and showed a specific detection of all cell types (> 0.90 pearson correlation). To test for significant differences in estimated RNA abundance between clusters, one-way ANOVA with Tukey's Multiple Comparison test was used.

Mice and EAE induction

$CD4^{Cre}$ mice *(91)* and B6.129S(FVB)-*Bcl6*$^{tm1.1Dent}$/J (named $Bcl6^{flox}$ or $Bcl6^{fl/fl}$) mice *(54)* were purchased from the Jackson laboratories. The $CD4^{Cre}Bcl6^{flox}$ strain was maintained by breeding the $Bcl6^{flox}$ allele to homozygosity (i.e. $Bcl6^{fl/fl}$) and breeding the Cre alleles in heterozygous to wildtype matings. Genotyping was done by routine PCR from ear punch DNA. All animal experiments were approved by the responsible state authorities (LANUV NRW) under reference number 84-02.04.2015.A319 and were performed in accordance with local regulations. Mice of both sexes (8-14 weeks old) were immunized s.c. in the flanks with an emulsion containing the myelin oligodendrocyte glycoprotein (MOG) peptide $MOG_{35–55}$ (150 µg/mouse) (GL Biochem (Shanghai) Ltd) and *M. tuberculosis* H37Ra extract (5 mg/ml, BD) in CFA (200 µl/mouse). Pertussis toxin (250 ng/mouse,

Sigma) was administered intraperitoneally on days 0 and 2. Mice were monitored daily and assigned grades for clinical signs of EAE using the following scoring system: 0, healthy; 1, paralyzed tail tip; 2, paralyzed tail; 3, waddling; 4, hind legs drag on the ground; 5, butt on the ground; 6, one paralyzed hind leg; 7, both paralyzed hind legs; 8, one paralyzed front leg (criterium to stop EAE); 9, both paralyzed front legs: 10, moribund or death. Detailed refinement procedures were performed according to the impairments of the mice. Mice with a score of >7 were euthanized. GraphPad Prism 5 was used for statistical analysis of all mouse-related data.

Isolation of CNS-infiltrating mononuclear cells

Mice were intracardially perfused with cold PBS under ketamin/xylazin anasthesia. The forebrain and cerebellum were dissected and spinal cords flushed out from the spinal canal with hydrostatic pressure. CNS tissue was cut into pieces and digested with collagenase D (2.5 mg/ml, Roche Diagnostics) and DNase I (0.05 mg/ml, Sigma) at 37 °C for 20 min. Mononuclear cells were isolated by passing the tissue through a 70 µm cell strainer, followed by a 70%/37% percoll gradient centrifugation. The interphase was removed, washed and re-suspended in culture medium containing 20 ng/ml PMA, 500 ng/ml ionomycin, GolgiStop, GolgiPlug (BD, each 1:1000 diluted). After 4 hours of incubation at 37 °C, cells were stained at RT for 30 min with anti-mouse antibodies (Biolegend, clones indicated): CD3 (17A2), CD4 (RM4-5 or GK1.5), B220 (RA3-6B2) and live/dead staining “Zombie NIR” (BD; 1:500) in PBS. Cells were fixed with the BD Cytofix/Cytoperm kit according to manufacturer instructions and stained with IL-17A (eBiosciences, eBio17B7) and IFNγ (BD, XMG1.2) each 1:100 diluted at 4°C for 30 min. Cells were washed and analysed using a Gallios flow cytometer (Beckman Coulter) and analysed using FlowJo V10.

Histology

For histology, mice were intracardially perfused with 20 ml cold PBS under ketamin/xylazin anasthesia and fixed by perfusion with 10 ml of 4 % paraformaldehyde (PFA). Spinal cord and spleen were removed and kept in PFA for 48 hours at 4 °C. The fixed spinal cords were cut into 3 mm thick transverse segments and embedded in paraffin. To evaluate demyelination, spinal cord sections were

816 stained with Luxol Fast Blue (LFB) and subsequently incubated with Periodic acid-Schiff (PAS).
817 Immunohistochemistry was performed using the biotin-streptavidin peroxidase technique (K5001,
818 Dako) in an immunostainer (AutostainerLink 48, Dako). Sections were pre-treated in a steamer
819 (treatment solutions pH 6.0 or pH 9.0 (Dako)) before incubation with the primary antibodies against
820 CD3 (clone CD3-12, BioRad, 1:100) or Mac3 (clone M3/84, BD, 1:100) or B220 (clone RA3-6B2, BD,
821 1:200). DAB was used as a chromogen. Stained sections were analysed with a keyence microscope and
822 pictures were taken with an Axioplot camara. ImageJ v1.48 was used to manually count infiltrated cells
823 and measure areas.
824

Supplementary Materials

852 **References**

853

854 1. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz,
855 M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells,
856 *Cell* **161**, 1187–1201 (2015).
857 2. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N.
858 Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A.
859 McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter
860 droplets, *Cell* **161**, 1202–1214 (2015).
861 3. J. Park, R. Shrestha, C. Qiu, A. Kondo, S. Huang, M. Werth, M. Li, J. Barasch, K. Suszták, Single-
862 cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease., *Science*
863 **360**, 758–763 (2018).
864 4. I. Tirosh, A. S. Venteicher, C. Hebert, L. E. Escalante, A. P. Patel, K. Yizhak, J. M. Fisher, C.
865 Rodman, C. Mount, M. G. Filbin, C. Neftel, N. Desai, J. Nyman, B. Izar, C. C. Luo, J. M. Francis, A.
866 A. Patel, M. L. Onozato, N. Riggi, K. J. Livak, D. Gennert, R. Satija, B. V. Nahed, W. T. Curry, R. L.
867 Martuza, R. Mylvaganam, A. J. Iafrate, M. P. Frosch, T. R. Golub, M. N. Rivera, G. Getz, O.
868 Rozenblatt-Rosen, D. P. Cahill, M. Monje, B. E. Bernstein, D. N. Louis, A. Regev, M. L. Suvà,
869 Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma, *Nature* **539**,
870 309–313 (2016).
871 5. E. Azizi, A. J. Carr, G. Plitas, A. E. Cornish, C. Konopacki, S. Prabhakaran, J. Nainys, K. Wu, V.
872 Kiseliovas, M. Setty, K. Choi, R. M. Fromme, P. Dao, P. T. McKenney, R. C. Wasti, K. Kadaveru, L.
873 Mazutis, A. Y. Rudensky, D. Pe'er, Single-Cell Map of Diverse Immune Phenotypes in the Breast
874 Tumor Microenvironment, *Cell* (2018), doi:10.1016/J.CELL.2018.05.060.
875 6. A. Giladi, I. Amit, Single-Cell Genomics: A Stepping Stone for Future Immunology Discoveries,
876 *Cell* **172**, 14–21 (2018).
877 7. A. K. Shalek, M. Benson, Single-cell analyses to tailor treatments., *Sci. Transl. Med.* **9**, eaan4730
878 (2017).
879 8. M. J. T. Stubbington, O. Rozenblatt-Rosen, A. Regev, S. A. Teichmann, Single-cell transcriptomics

to explore the immune system in health and disease., *Science* **358**, 58–63 (2017).

9. M. G. Filbin, I. Tirosh, V. Hovestadt, M. L. Shaw, L. E. Escalante, N. D. Mathewson, C. Neftel, N. Frank, K. Pelton, C. M. Hebert, C. Haberler, K. Yizhak, J. Gojo, K. Egervari, C. Mount, P. van Galen, D. M. Bonal, Q.-D. Nguyen, A. Beck, C. Sinai, T. Czech, C. Dorfer, L. Goumnerova, C. Lavarino, A. M. Carcaboso, J. Mora, R. Mylvaganam, C. C. Luo, A. Peyrl, M. Popović, A. Azizi, T. T. Batchelor, M. P. Frosch, M. Martinez-Lage, M. W. Kieran, P. Bandopadhayay, R. Beroukhim, G. Fritsch, G. Getz, O. Rozenblatt-Rosen, K. W. Wucherpfennig, D. N. Louis, M. Monje, I. Slavc, K. L. Ligon, T. R. Golub, A. Regev, B. E. Bernstein, M. L. Suvà, Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq., *Science* **360**, 331–335 (2018).

10. H. Zhang, C. A. A. Lee, Z. Li, J. R. Garbe, C. R. Eide, R. Petegrosso, R. Kuang, J. Tolar, A. Khan, Ed. A multitask clustering approach for single-cell RNA-seq analysis in Recessive Dystrophic Epidermolysis Bullosa, *PLOS Comput. Biol.* **14**, e1006053 (2018).

11. E. Der, S. Ranabothu, H. Suryawanshi, K. M. Akat, R. Clancy, P. Morozov, M. Kustagi, M. Czuppa, P. Izmirly, H. M. Belmont, T. Wang, N. Jordan, N. Bornkamp, J. Nwaukoni, J. Martinez, B. Goilav, J. P. Buyon, T. Tuschl, C. Putterman, Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis, *JCI Insight* **2** (2017), doi:10.1172/JCI.INSIGHT.93009.

12. K. D. Korthauer, L.-F. Chu, M. A. Newton, Y. Li, J. Thomson, R. Stewart, C. Kendziorski, A statistical approach for identifying differential distributions in single-cell RNA-seq experiments, *Genome Biol.* **17**, 222 (2016).

13. A. Compston, A. Coles, Multiple sclerosis, *Lancet* **372**, 1502–1517 (2008).

14. H.-C. Von Büdingen, T. C. Kuo, M. Sirota, C. J. Van Belle, L. Apeltsin, J. Glanville, B. A. Cree, P. Gourraud, A. Schwartzburg, G. Huerta, D. Telman, P. D. Sundar, T. Casey, D. R. Cox, S. L. Hauser, B cell exchange across the blood-brain barrier in multiple sclerosis, *J. Clin. Invest.* **122**, 24–28 (2012).

15. S. L. Hauser, A. Bar-Or, G. Comi, G. Giovannoni, H.-P. Hartung, B. Hemmer, F. Lublin, X. Montalban, K. W. Rammohan, K. Selmaj, A. Traboulsee, J. S. Wolinsky, D. L. Arnold, G. Klingelschmitt, D. Masterman, P. Fontoura, S. Belachew, P. Chin, N. Mairon, H. Garren, L. Kappos, Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis, *N. Engl. J. Med.* **376**, 221–

234 (2017).

16. H. W. Kreth, R. Dunker, H. Rodt, R. Meyermann, Immunohistochemical identification of T-lymphocytes in the central nervous system of patients with multiple sclerosis and subacute sclerosing panencephalitis, *J. Neuroimmunol.* **2**, 177–183 (1982).

17. J. Machado-Santos, E. Saji, A. R. Tröscher, M. Paunovic, R. Liblau, G. Gabriely, C. G. Bien, J. Bauer, H. Lassmann, The compartmentalized inflammatory response in the multiple sclerosis brain is composed of tissue-resident CD8+ T lymphocytes and B cells, *Brain* , 2066–2082 (2018).

18. M. Rangachari, V. K. Kuchroo, Using EAE to better understand principles of immune function and autoimmune pathology, *J. Autoimmun.* **45**, 31–39 (2013).

19. B. Brynedal, M. Khademi, E. Wallström, J. Hillert, T. Olsson, K. Duvefelt, Gene expression profiling in multiple sclerosis: A disease of the central nervous system, but with relapses triggered in the periphery?, *Neurobiol. Dis.* **37**, 613–621 (2010).

20. J. J. Iliff, M. Wang, Y. Liao, B. A. Plogg, W. Peng, S. A. Goldman, E. A. Nagelhus, M. Nedergaard, A Paravascular Pathway Facilitates CSF Flow Through the Brain Parenchyma and the Clearance of Interstitial Solutes , Including Amyloid b, *Sci. Transl. Med.* **4**, 1–11 (2012).

21. C. Schläger, H. Körner, M. Krueger, S. Vidoli, M. Haberl, D. Mielke, E. Brylla, T. Issekutz, C. Cabanãs, P. J. Nelson, T. Ziemssen, V. Rohde, I. Bechmann, D. Lodygin, F. Odoardi, A. Flügel, Effector T-cell trafficking between the leptomeninges and the cerebrospinal fluid, *Nature* **530**, 349–353 (2016).

22. B. Engelhardt, R. O. Carare, I. Bechmann, A. Flügel, J. D. Laman, R. O. Weller, Vascular, glial, and lymphatic immune gateways of the central nervous system, *Acta Neuropathol.* **132**, 317–338 (2016).

23. S. Han, Y. C. Lin, T. Wu, A. D. Salgado, I. Mexhitaj, S. C. Wuest, E. Romm, J. Ohayon, R. Goldbach-Mansky, A. Vanderver, A. Marques, C. Toro, P. Williamson, I. Cortese, B. Bielekova, Comprehensive Immunophenotyping of Cerebrospinal Fluid Cells in Patients with Neuroimmunological Diseases, *J. Immunol.* **192**, 2551–2563 (2014).

24. S. M. Brändle, B. Obermeier, M. Senel, J. Bruder, R. Mentele, M. Khademi, T. Olsson, H. Tumani, W. Kristoferitsch, F. Lottspeich, H. Wekerle, R. Hohlfeld, K. Dornmair, Distinct oligoclonal

band antibodies in multiple sclerosis recognize ubiquitous self-proteins, *Proc. Natl. Acad. Sci.* **113**, 7864–7869 (2016).

25. S. Cepok, M. Jacobsen, S. Schock, B. Omer, S. Jaekel, I. Böddeker, W. H. Oertel, N. Sommer, B. Hemmer, Patterns of cerebrospinal fluid pathology correlate with disease progression in multiple sclerosis., *Brain* **124**, 2169–76 (2001).

26. M. C. Kowarik, V. Grummel, S. Wemlinger, D. Buck, M. S. Weber, A. Berthele, B. Hemmer, Immune cell subtyping in the cerebrospinal fluid of patients with neurological diseases, *J. Neurol.* **261**, 130–143 (2014).

27. A. Corcione, S. Casazza, E. Ferretti, D. Giunti, E. Zappia, A. Pistorio, C. Gambini, G. L. Mancardi, A. Uccelli, V. Pistoia, Recapitulation of B cell differentiation in the central nervous system of patients with multiple sclerosis., *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11064–9 (2004).

28. J. Haas, I. Bekeredjian-Ding, M. Milkova, B. Balint, A. Schwarz, M. Korporal, S. Jarius, B. Fritz, H. M. Lorenz, B. Wildemann, B cells undergo unique compartmentalized redistribution in multiple sclerosis, *J. Autoimmun.* **37**, 289–299 (2011).

29. R. W. Steele, D. J. Marmer, M. D. O'Brien, S. T. Tyson, C. R. Steele, Leukocyte survival in cerebrospinal fluid., *J. Clin. Microbiol.* **23**, 965–6 (1986).

30. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells, *Nat. Commun.* **8** (2017), doi:10.1038/ncomms14049.

31. D. E. Berezovsky, B. B. Bruce, C. Vasseneix, J. H. Peragallo, N. J. Newman, V. Biousse, Cerebrospinal fluid total protein in idiopathic intracranial hypertension, *J. Neurol. Sci.* **381**, 226–229 (2017).

32. M. Wall, Idiopathic intracranial hypertension., *Neurol Clin.* **28**, 593–617 (2010).

33. A. Waschbisch, S. Schröder, D. Schraudner, L. Sammet, B. Weksler, A. Melms, S. Pfeifenbring, C. Stadelmann, S. Schwab, R. A. Linker, Pivotal Role for CD16+ Monocytes in Immune Surveillance

964 of the Central Nervous System., *J. Immunol.* **196**, 1558–67 (2016).

965 34. E. Rodríguez-Martín, C. Picón, L. Costa-Frossard, R. Alenda, S. Sainz de la Maza, E. Roldán, M.

966 Espiño, L. M. Villar, J. C. Álvarez-Cermeño, Natural killer cell subsets in cerebrospinal fluid of

967 patients with multiple sclerosis, *Clin. Exp. Immunol.* **180**, 243–249 (2015).

968 35. E. L. Eggers, B. a. Michel, H. Wu, S. Wang, C. J. Bevan, A. Abounasr, N. S. Pierson, A. Bischof,

969 M. Kazer, E. Leitner, A. L. Greenfield, S. Demuth, M. R. Wilson, R. G. Henry, B. a. C. Cree, S. L.

970 Hauser, H.-C. von Büdingen, Clonal relationships of CSF B cells in treatment-naive multiple sclerosis

971 patients, *JCI Insight* **2** (2017), doi:10.1172/jci.insight.92724.

972 36. J. T. Gaublomme, N. Yosef, Y. Lee, R. S. Gertner, L. V Yang, C. Wu, Single-Cell Genomics

973 Unveils Critical Regulators of Th17 Cell Pathogenicity, (2015).

974 37. D. DeTomaso, N. Yosef, FastProject: a tool for low-dimensional analysis of single-cell RNA-Seq

975 data, *BMC Bioinformatics* **17**, 315 (2016).

976 38. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn,

977 A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles., *Nat. Methods* **12**,

978 453–7 (2015).

979 39. S. Sawcer, G. Hellenthal, M. Pirinen, C. C. A. Spencer, N. A. Patsopoulos, L. Moutsianas, A.

980 Dilthey, Z. Su, C. Freeman, S. E. Hunt, S. Edkins, E. Gray, D. R. Booth, S. C. Potter, A. Goris, G.

981 Band, A. Bang Oturai, A. Strange, J. Saarela, C. Bellenguez, B. Fontaine, M. Gillman, B. Hemmer, R.

982 Gwilliam, F. Zipp, A. Jayakumar, R. Martin, S. Leslie, S. Hawkins, E. Giannoulatou, S. D'alfonso, H.

983 Blackburn, F. Martinelli Boneschi, J. Liddle, H. F. Harbo, M. L. Perez, A. Spurkland, M. J. Waller,

984 M. P. Mycko, M. Ricketts, M. Comabella, N. Hammond, I. Kockum, O. T. McCann, M. Ban, P.

985 Whittaker, A. Kemppinen, P. Weston, C. Hawkins, S. Widaa, J. Zajicek, S. Dronov, N. Robertson, S.

986 J. Bumpstead, L. F. Barcellos, R. Ravindrarajah, R. Abraham, L. Alfredsson, K. Ardlie, C. Aubin, A.

987 Baker, K. Baker, S. E. Baranzini, L. Bergamaschi, R. Bergamaschi, A. Bernstein, A. Berthele, M.

988 Boggild, J. P. Bradfield, D. Brassat, S. A. Broadley, D. Buck, H. Butzkueven, R. Capra, W. M.

989 Carroll, P. Cavalla, E. G. Celius, S. Cepok, R. Chiavacci, F. Clerget-Darpoux, K. Clysters, G. Comi,

990 M. Cossburn, I. Cournu-Rebeix, M. B. Cox, W. Cozen, B. A. C. Cree, A. H. Cross, D. Cusi, M. J.

991 Daly, E. Davis, P. I. W. de Bakker, M. Debouverie, M. B. D'hooghe, K. Dixon, R. Dobosi, B. Dubois,

992 D. Ellinghaus, I. Elovaara, F. Esposito, C. Fontenille, S. Foote, A. Franke, D. Galimberti, A. Ghezzi,
993 J. Glessner, R. Gomez, O. Gout, C. Graham, S. F. A. Grant, F. Rosa Guerini, H. Hakonarson, P. Hall,
994 A. Hamsten, H.-P. Hartung, R. N. Heard, S. Heath, J. Hobart, M. Hoshi, C. Infante-Duarte, G.
995 Ingram, W. Ingram, T. Islam, M. Jagodic, M. Kabesch, A. G. Kermode, T. J. Kilpatrick, C. Kim, N.
996 Klopp, K. Koivisto, M. Larsson, M. Lathrop, J. S. Lechner-Scott, M. A. Leone, V. Leppä, U.
997 Liljedahl, I. Lima Bomfim, R. R. Lincoln, J. Link, J. Liu, Å. R. Lorentzen, S. Lupoli, F. Macciardi, T.
998 Mack, M. Marriott, V. Martinelli, D. Mason, J. L. McCauley, F. Mentch, I.-L. Mero, T. Mihalova, X.
999 Montalban, J. Mottershead, K.-M. Myhr, P. Naldi, W. Ollier, A. Page, A. Palotie, J. Pelletier, L.
1000 Piccio, T. Pickersgill, F. Piehl, S. Pobywajlo, H. L. Quach, P. P. Ramsay, M. Reunanen, R. Reynolds,
1001 J. D. Rioux, M. Rodegher, S. Roesner, J. P. Rubio, I.-M. Rückert, M. Salvetti, E. Salvi, A.
1002 Santaniello, C. A. Schaefer, S. Schreiber, C. Schulze, R. J. Scott, F. Sellebjerg, K. W. Selmaj, D.
1003 Sexton, L. Shen, B. Simms-Acuna, S. Skidmore, P. M. A. Sleiman, C. Smestad, P. S. Sørensen, H. B.
1004 Søndergaard, J. Stankovich, R. C. Strange, A.-M. Sulonen, E. Sundqvist, A.-C. Syvänen, F. Taddeo,
1005 B. Taylor, J. M. Blackwell, P. Tienari, E. Bramon, A. Tourbah, M. A. Brown, E. Tronczynska, J. P.
1006 Casas, N. Tubridy, A. Corvin, J. Vickery, J. Jankowski, P. Villoslada, H. S. Markus, K. Wang, C. G.
1007 Mathew, J. Wason, C. N. A. Palmer, H.-E. Wichmann, R. Plomin, E. Willoughby, A. Rautanen, J.
1008 Winkelmann, M. Wittig, R. C. Trembath, J. Yaouanq, A. C. Viswanathan, H. Zhang, N. W. Wood, R.
1009 Zuvich, P. Deloukas, C. Langford, A. Duncanson, J. R. Oksenberg, M. A. Pericak-Vance, J. L.
1010 Haines, T. Olsson, J. Hillert, A. J. Ivinson, P. L. De Jager, L. Peltonen, G. J. Stewart, D. A. Hafler, S.
1011 L. Hauser, G. McVean, P. Donnelly, A. Compston, Genetic risk and a primary role for cell-mediated
1012 immune mechanisms in multiple sclerosis, *Nature* **476**, 214–219 (2011).
1013 40. P. L. De Jager, I. M. S. G. Consortium, The Multiple Sclerosis Genomic Map: Role of peripheral
1014 immune cells and resident microglia in susceptibility, , 1–43 (2017).
1015 41. a. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. a. Gillette, a.
1016 Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A
1017 knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci.*
1018 **102**, 15545–15550 (2005).
1019 42. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov,

Molecular signatures database (MSigDB) 3.0., *Bioinformatics* **27**, 1739–40 (2011).

43. K. Kandasamy, S. Mohan, R. Raju, S. Keerthikumar, G. S. S. Kumar, A. K. Venugopal, D. Telikicherla, D. J. Navarro, S. Mathivanan, C. Pecquet, S. K. Gollapudi, S. G. Tattikota, S. Mohan, H. Padhukasahasram, Y. Subbannayya, R. Goel, H. K. C. Jacob, J. Zhong, R. Sekhar, V. Nanjappa, L. Balakrishnan, R. Subbaiah, Y. l Ramachandra, A. Rahiman, T. s Keshava Prasad, J.-X. Lin, J. C. D. Houtman, S. Desiderio, J.-C. Renauld, S. Constantinescu, O. Ohara, T. Hirano, M. Kubo, S. Singh, P. Khatri, S. Draghici, G. D. Bader, C. Sander, W. J. Leonard, A. Pandey, NetPath: a public resource of curated signal transduction pathways, *Genome Biol.* **11**, R3 (2010).

44. A. Crawford, J. M. Angelosanto, C. Kao, T. A. Doering, P. M. Odorizzi, B. E. Barnett, E. J. Wherry, Molecular and transcriptional basis of $CD4^+$ T cell dysfunction during chronic infection., *Immunity* **40**, 289–302 (2014).

45. T. C. Wirth, H.-H. Xue, D. Rai, J. T. Sabel, T. Bair, J. T. Harty, V. P. Badovinac, Repetitive antigen stimulation induces stepwise transcriptome diversification but preserves a core signature of memory CD8(+) T cell differentiation., *Immunity* **33**, 128–40 (2010).

46. T. Lönnberg, V. Svensson, K. R. James, D. Fernandez-Ruiz, I. Sebina, R. Montandon, M. S. F. Soon, L. G. Fogg, A. S. Nair, U. N. Liligeto, M. J. T. Stubbington, L.-H. Ly, F. O. Bagger, M. Zwiessele, N. D. Lawrence, F. Souza-Fonseca-Guimaraes, P. T. Bunn, C. R. Engwerda, W. R. Heath, O. Billker, O. Stegle, A. Haque, S. A. Teichmann, Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves $T_H 1/T_{FH}$ fate bifurcation in malaria, *Sci. Immunol.* **2**, eaal2192 (2017).

47. S. Xiao, N. Yosef, J. Yang, Y. Wang, L. Zhou, C. Zhu, C. Wu, E. Baloglu, D. Schmidt, R. Ramesh, M. Lobera, M. S. Sundrud, P.-Y. Tsai, Z. Xiang, J. Wang, Y. Xu, X. Lin, K. Kretschmer, P. B. Rahl, R. A. Young, Z. Zhong, D. A. Hafler, A. Regev, S. Ghosh, A. Marson, V. K. Kuchroo, Small-molecule RORγt antagonists inhibit T helper 17 cell transcriptional network by divergent mechanisms., *Immunity* **40**, 477–89 (2014).

48. X. Liu, X. Chen, B. Zhong, A. Wang, X. Wang, F. Chu, R. I. Nurieva, X. Yan, P. Chen, L. G. van der Flier, H. Nakatsukasa, S. S. Neelapu, W. Chen, H. Clevers, Q. Tian, H. Qi, L. Wei, C. Dong, Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development, *Nature*

507, 513–518 (2014).

49. S. Crotty, Follicular Helper CD4 T Cells (T_{FH}), *Annu. Rev. Immunol.* **29**, 621–663 (2011).

50. R. I. Nurieva, Y. Chung, G. J. Martinez, X. O. Yang, S. Tanaka, T. D. Matskevitch, Y. Wang, C. Dong, Bcl6 mediates the development of T follicular helper cells, *Science (80-.).* **325**, 1001–1005 (2010).

51. X. Fan, T. Jin, S. Zhao, C. Liu, J. Han, X. Jiang, Y. Jiang, Circulating CCR7+ICOS+ memory T follicular helper cells in patients with multiple sclerosis, *PLoS One* **10**, 1–14 (2015).

52. J. R. Christensen, L. Börnsen, R. Ratzer, F. Piehl, M. Khademi, T. Olsson, P. S. Sørensen, F. Sellebjerg, Systemic Inflammation in Progressive Multiple Sclerosis Involves Follicular T-Helper, Th17- and Activated B-Cells and Correlates with Progression, *PLoS One* **8**, 1–11 (2013).

53. D. a. Rao, M. F. Gurish, J. L. Marshall, K. Slowikowski, C. Y. Fonseka, Y. Liu, L. T. Donlin, L. a. Henderson, K. Wei, F. Mizoguchi, N. C. Teslovich, M. E. Weinblatt, E. M. Massarotti, J. S. Coblyn, S. M. Helfgott, Y. C. Lee, D. J. Todd, V. P. Bykerk, S. M. Goodman, A. B. Pernis, L. B. Ivashkiv, E. W. Karlson, P. a. Nigrovic, A. Filer, C. D. Buckley, J. a. Lederer, S. Raychaudhuri, M. B. Brenner, Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis, *Nature* **542**, 110–114 (2017).

54. K. Hollister, S. Kusam, H. Wu, N. Clegg, a. Mondal, D. V. Sawant, a. L. Dent, Insights into the Role of Bcl6 in Follicular Th Cells Using a New Conditional Mutant Mouse Model, *J. Immunol.* **191**, 3705–3711 (2013).

55. Y. Enose-Akahata, S. Azodi, B. R. Smith, B. J. Billioux, A. Vellucci, N. Ngouth, Y. Tanaka, J. Ohayon, I. Cortese, A. Nath, S. Jacobson, C. R. M. Bangham, Ed. Immunophenotypic characterization of CSF B cells in virus-associated neuroinflammatory diseases, *PLOS Pathog.* **14**, e1007042 (2018).

56. J. Guo, C. Zhao, F. Wu, L. Tao, C. Zhang, D. Zhao, S. Yang, D. Jiang, J. Wang, Y. Sun, Z. Li, H. Li, K. Yang, T Follicular Helper-Like Cells Are Involved in the Pathogenesis of Experimental Autoimmune Encephalomyelitis, *Front. Immunol.* **9**, 944 (2018).

57. A. Lossius, A. Tomescu-Baciu, T. Holmøy, C. A. Vedeler, E. Røsjø, Å. R. Lorentzen, I. Casetta, F. Vartdal, Selective intrathecal enrichment of G1m1-positive B cells in multiple sclerosis., *Ann. Clin.*

Transl. Neurol. **4**, 756–761 (2017).

58. S. Cepok, B. Rosche, V. Grummel, F. Vogel, D. Zhou, J. Sayn, N. Sommer, H.-P. Hartung, B. Hemmer, Short-lived plasma blasts are the main B cell effector subset during the course of multiple sclerosis, *Brain* **128**, 1667–1676 (2005).

59. C. M. Lill, B. M. M. Schjeide, C. Graetz, M. Ban, A. Alcina, M. a. Ortiz, J. Pérez, V. Damotte, D. Booth, A. Lopez De Lapuente, L. Broer, M. Schilling, D. a. Akkad, O. Aktas, I. Alloza, A. Antigüedad, R. Arroyo, P. Blaschke, M. Buttmann, A. Chan, A. Compston, I. Cournu-Rebeix, T. Dörner, J. T. Epplen, Ó. Fernández, L. A. Gerdes, L. Guillot-Noël, H. P. Hartung, S. Hoffjan, G. Izquierdo, A. Kemppinen, A. Kroner, C. Kubisch, T. Kümpfel, S. C. Li, U. Lindenberger, P. Lohse, C. Lubetzki, F. Luessi, S. Malhotra, J. Mescheriakova, X. Montalban, C. Papeix, L. F. Paredes, P. Rieckmann, E. Steinhagen-Thiessen, A. Winkelmann, U. K. Zettl, R. Hintzen, K. Vandenbroeck, G. Stewart, B. Fontaine, M. Comabella, E. Urcelay, F. Matesanz, S. Sawcer, L. Bertram, F. Zipp, MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis, *Brain* **136**, 1778–1782 (2013).

60. F. Humby, M. Bombardieri, A. Manzo, S. Kelly, M. C. Blades, B. Kirkham, J. Spencer, C. Pitzalis, Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium., *PLoS Med.* **6**, e1 (2009).

61. R. Magliozzi, O. Howell, A. Vora, B. Serafini, R. Nicholas, M. Puopolo, R. Reynolds, F. Aloisi, Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology, *Brain* **130**, 1089–1104 (2006).

62. B. Serafini, B. Rosicarelli, R. Magliozzi, E. Stigliano, F. Aloisi, Detection of Ectopic B-cell Follicles with Germinal Centers in the Meninges of Patients with Secondary Progressive Multiple Sclerosis, *Brain Pathol.* **14**, 164–174 (2004).

63. S. R. Slight, J. Rangel-Moreno, R. Gopal, Y. Lin, B. A. Fallert Junecko, S. Mehra, M. Selman, E. Becerril-Villanueva, J. Baquera-Heredia, L. Pavon, D. Kaushal, T. A. Reinhart, T. D. Randall, S. A. Khader, CXCR5^{+} T helper cells mediate protective immunity against tuberculosis., *J. Clin. Invest.* **123**, 712–26 (2013).

64. G. Androdias, R. Reynolds, M. Chanal, C. Ritleng, C. Confavreux, S. Nataf, Meningeal T cells

associate with diffuse axonal loss in multiple sclerosis spinal cords, *Ann. Neurol.* **68**, 465–476 (2010).

65. S. R. Choi, O. W. Howell, D. Carassiti, R. Magliozzi, D. Gveric, P. A. Muraro, R. Nicholas, F. Roncaroli, R. Reynolds, Meningeal inflammation plays a role in the pathology of primary progressive multiple sclerosis, *Brain* **135**, 2925–2937 (2012).

66. R. Magliozzi, O. W. Howell, C. Reeves, F. Roncaroli, R. Nicholas, B. Serafini, F. Aloisi, R. Reynolds, A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis, *Ann. Neurol.* **68**, 477–493 (2010).

67. C. Teunissen, T. Menge, A. Altintas, J. C. Álvarez-Cermeño, A. Bertolotto, F. S. Berven, L. Brundin, M. Comabella, M. Degn, F. Deisenhammer, F. Fazekas, D. Franciotta, J. L. Frederiksen, D. Galimberti, S. Gnanapavan, H. Hegen, B. Hemmer, R. Hintzen, S. Hughes, E. Iacobaeus, A. C. Kroksveen, J. Kuhle, J. Richert, H. Tumani, L. M. Villar, J. Drulovic, I. Dujmovic, M. Khalil, A. Bartos, Consensus definitions and application guidelines for control groups in cerebrospinal fluid biomarker studies in multiple sclerosis, *Mult. Scler. J.* **19**, 1802–1809 (2013).

68. G. M. Swanson, J. C. Bailar, Selection and description of cancer clinical trials participants?Science or happenstance?, *Cancer* **95**, 950–959 (2002).

69. D. L. Vargas, W. R. Tyor, Update on disease-modifying therapies for multiple sclerosis., *J. Investig. Med.* **65**, 883–891 (2017).

70. M. K. Singh, T. F. Scott, W. A. LaFramboise, F. Z. Hu, J. C. Post, G. D. Ehrlich, Gene expression changes in peripheral blood mononuclear cells from multiple sclerosis patients undergoing β-interferon therapy, *J. Neurol. Sci.* **258**, 52–59 (2007).

71. A. Achiron, M. Gurevich, N. Friedman, N. Kaminski, M. Mandel, Blood transcriptional signatures of multiple sclerosis: Unique gene expression of disease activity, *Ann. Neurol.* **55**, 410–417 (2004).

72. J. Satoh, M. Nakanishi, F. Koike, H. Onoue, T. Aranami, T. Yamamoto, M. Kawai, S. Kikuchi, K. Nomura, K. Yokoyama, K. Ota, T. Saito, M. Ohta, S. Miyake, T. Kanda, T. Fukazawa, T. Yamamura, T cell gene expression profiling identifies distinct subgroups of Japanese multiple sclerosis patients., *J. Neuroimmunol.* **174**, 108–18 (2006).

73. R. Menon, M. Di Dario, C. Cordiglieri, S. Musio, L. La Mantia, C. Milanese, A. L. Di Stefano, M. Crabbio, D. Franciotta, R. Bergamaschi, R. Pedotti, E. Medico, C. Farina, Gender-based blood

1132 transcriptomes and interactomes in multiple sclerosis: Involvement of SP1 dependent gene
1133 transcription, *J. Autoimmun.* **38**, J144–J155 (2012).
1134 74. K. S. Gandhi, F. C. McKay, M. Cox, C. Riveros, N. Armstrong, R. N. Heard, S. Vucic, D. W.
1135 Williams, J. Stankovich, M. Brown, P. Danoy, G. J. Stewart, S. Broadley, P. Moscato, J. Lechner-
1136 Scott, R. J. Scott, D. R. Booth, L. Griffiths, M. Slee, S. Browning, B. L. Browning, T. Kilpatrick, J.
1137 Rubio, V. Perreau, H. Butzkeuven, M. Tanner, J. Wiley, S. Foote, J. Stankovich, B. Taylor, A.
1138 Kermode, B. Carroll, M. Bahlo, The multiple sclerosis whole blood mRNA transcriptome and genetic
1139 associations indicate dysregulation of specific T cell pathways in pathogenesis, *Hum. Mol. Genet.* **19**,
1140 2134–2143 (2010).
1141 75. L. Ottoboni, B. T. Keenan, P. Tamayo, M. Kuchroo, J. P. Mesirov, G. J. Buckle, S. J. Khoury, D.
1142 A. Hafler, H. L. Weiner, P. L. De Jager, An RNA profile identifies two subsets of multiple sclerosis
1143 patients differing in disease activity., *Sci. Transl. Med.* **4**, 153ra131 (2012).
1144 76. S. Srinivasan, M. Di Dario, A. Russo, R. Menon, E. Brini, M. Romeo, F. Sangalli, G. D. Costa, M.
1145 Rodegher, M. Radaelli, L. Moiola, D. Cantarella, E. Medico, G. Martino, R. Furlan, V. Martinelli, G.
1146 Comi, C. Farina, Dysregulation of MS risk genes and pathways at distinct stages of disease., *Neurol.*
1147 *Neuroimmunol. neuroinflammation* **4**, e337 (2017).
1148 77. D. Nickles, H. P. Chen, M. M. Li, P. Khankhanian, L. Madireddy, S. J. Caillier, A. Santaniello, B.
1149 A. C. Cree, D. Pelletier, S. L. Hauser, J. R. Oksenberg, S. E. Baranzini, Blood RNA profiling in a
1150 large cohort of multiple sclerosis patients and healthy controls, *Hum. Mol. Genet.* **22**, 4194–4205
1151 (2013).
1152 78. Y. Cao, B. A. Goods, K. Raddassi, G. T. Nepom, W. W. Kwok, J. C. Love, D. A. Hafler,
1153 Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple
1154 sclerosis., *Sci. Transl. Med.* **7**, 287ra74 (2015).
1155 79. M. Filippi, M. A. Rocca, O. Ciccarelli, N. De Stefano, N. Evangelou, L. Kappos, A. Rovira, J.
1156 Sastre-Garriga, M. Tintorè, J. L. Frederiksen, C. Gasperini, J. Palace, D. S. Reich, B. Banwell, X.
1157 Montalban, F. Barkhof, MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus
1158 guidelines, *Lancet Neurol.* **15**, 292–303 (2016).
1159 80. X. Montalban, M. Tintoré, J. Swanton, F. Barkhof, F. Fazekas, M. Filippi, J. Frederiksen, L.

Kappos, J. Palace, C. Polman, M. Rovaris, N. de Stefano, A. Thompson, T. Yousry, A. Rovira, D. H. Miller, MRI criteria for MS in patients with clinically isolated syndromes., *Neurology* **74**, 427–34 (2010).

81. M. B. Cole, D. Risso, A. Wagner, D. DeTomaso, J. Ngai, E. Purdom, S. Dudoit, N. Yosef, Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq, *bioRxiv* , 235382 (2017).

82. E. Eisenberg, E. Y. Levanon, Human housekeeping genes are compact, *Trends Genet.* **19**, 362–365 (2003).

83. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples, *Nat. Biotechnol.* **32**, 896–902 (2014).

84. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data, *Nat. Biotechnol.* **33**, 495–502 (2015).

85. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: precision weights unlock linear model analysis tools for RNA-seq read counts, *Genome Biol.* **15**, R29 (2014).

86. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments, *Ann. Appl. Stat.* (2011), doi:10.1214/11-AOAS466.

87. E. Martin-Gayo, M. B. Cole, K. E. Kolb, Z. Ouyang, J. Cronin, S. W. Kazer, J. Ordovas-Montanes, M. Lichterfeld, B. D. Walker, N. Yosef, A. K. Shalek, X. G. Yu, A Reproducibility-Based Computational Framework Identifies an Inducible, Enhanced Antiviral State in Dendritic Cells from HIV-1 Elite Controllers., *Genome Biol.* **19**, 10 (2018).

88. S. Picelli, O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2, *Nat. Protoc.* **9**, 171–181 (2014).

89. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.* **10**, R25 (2009).

90. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics* **12**, 323 (2011).

91. P. P. Lee, D. R. Fitzpatrick, C. Beard, H. K. Jessup, S. Lehar, K. W. Makar, M. Pérez-Melgosa, M. T. Sweetser, M. S. Schlissel, S. Nguyen, S. R. Cherry, J. H. Tsai, S. M. Tucker, W. M. Weaver, A.

Kelso, R. Jaenisch, C. B. Wilson, A Critical Role for Dnmt1 and DNA Methylation in T Cell Development, Function, and Survival, *Immunity* **15**, 763–774 (2001).

Acknowledgements

We thank Claudia Kemming, Anna-Lena Börsch, Maik Höfer, Gabriele Berens, and Kirsten Weiss for technical assistance.

Funding

G.M.z.H. was supported in part by grants from the Deutsche Forschungsgemeinschaft (DFG, grant number ME4050/4-1), from the Gemeinnützige Hertie Stiftung, from the Innovative Medical Research (IMF) program of the Westfälische Wilhelms-University Münster, and from the Ministerium für Innovation, Wissenschaft und Forschung (MIWF) des Landes Nordrhein-Westfalen. This project was funded in part by the Sonderforschungsbereich Transregio 128 of the DFG (to S.G.M, A09 to H.W. and C.C.G., Z02 to H.W. and T.K.).

Author contributions

D.S., M.H., T.L., J.W. performed experiments, M.C., K.B., N.Y. performed computational analyses, M.S., C.K. recruited patients and performed lumbar punctures, C.G. processed CSF samples, T.K. performed histology, S.G.M., H.W. co-supervised the study, N.Y., G.M.z.H. conceived and supervised the study and wrote the manuscript. All authors critically revised the manuscript.

Competing Interests

The authors declare no competing interests.

Figures

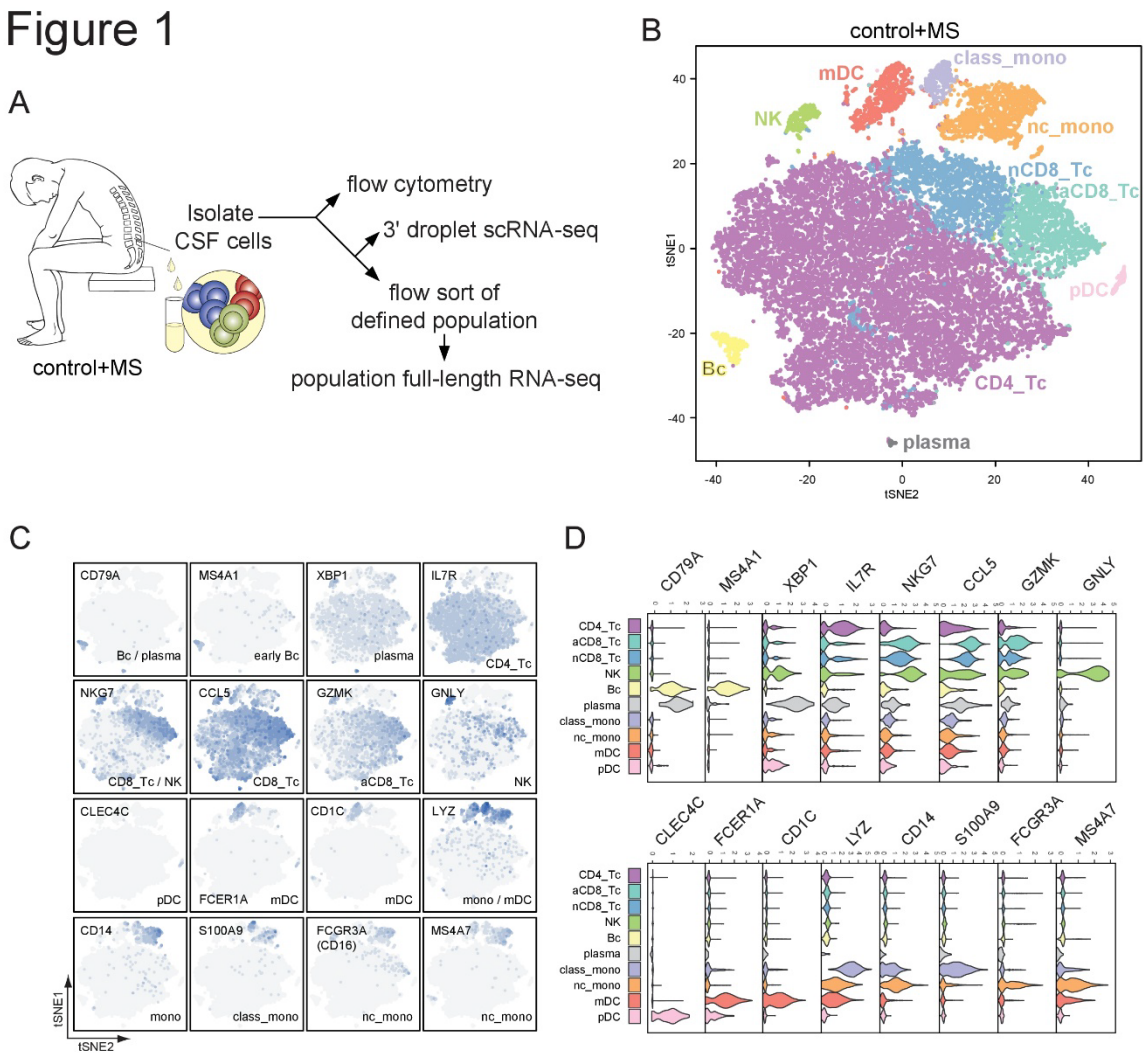


Fig. 1. Single-cell transcriptomics reconstructs the CSF leukocyte composition.

(A) Schematic of the study sampling and processing. CSF cells of all control (n = 22) and multiple sclerosis (MS) (n = 26) donors were analysed by flow cytometry. Subsequently, scRNA-seq of unsorted CSF cells (cohort 1, n = 6 donors each group) and bulk RNA-seq of sorted T follicular helper (TFH) cells (cohort 3, n = 9 donors each group) were performed on randomly selected donor samples. Cohort 2 was only used for flow cytometry and not processed for sequencing (B) t-distributed stochastic neighbour embedding (t-SNE) plot of 10 color-coded cell clusters identified by scRNA-seq after quality control filtering and normalization (Methods) in 22,357 total merged control- (n = 4) and MS-derived (n = 4) CSF cells. Cluster identity was manually assigned based on marker gene expression: (C) Feature plots, representing all 22,357 donor cells as in panel B, showing expression of selected marker genes,

differentially-expressed in one vs. all comparisons (Methods). Dark blue colours indicate high log-expression while light grey indicates non-expression. Selected protein names are provided for clarity, with expressing cell types indicated. (D) Stacked violin plots of the same marker genes in specified cell clusters. Cluster key: *CD4_Tc* CD4$^+$ T cells, *aCD8_Tc* / *nCD8_Tc* activated / naïve CD8$^+$ T cells, *NK* natural killer cells, *Bc* B cells, *plasma* plasma cells, *class_mono* / *nc_mono* classical / non-classical monocytes, *mDC* / *pDC* myeloid / plasmacytoid dendritic cells.

Figure 2

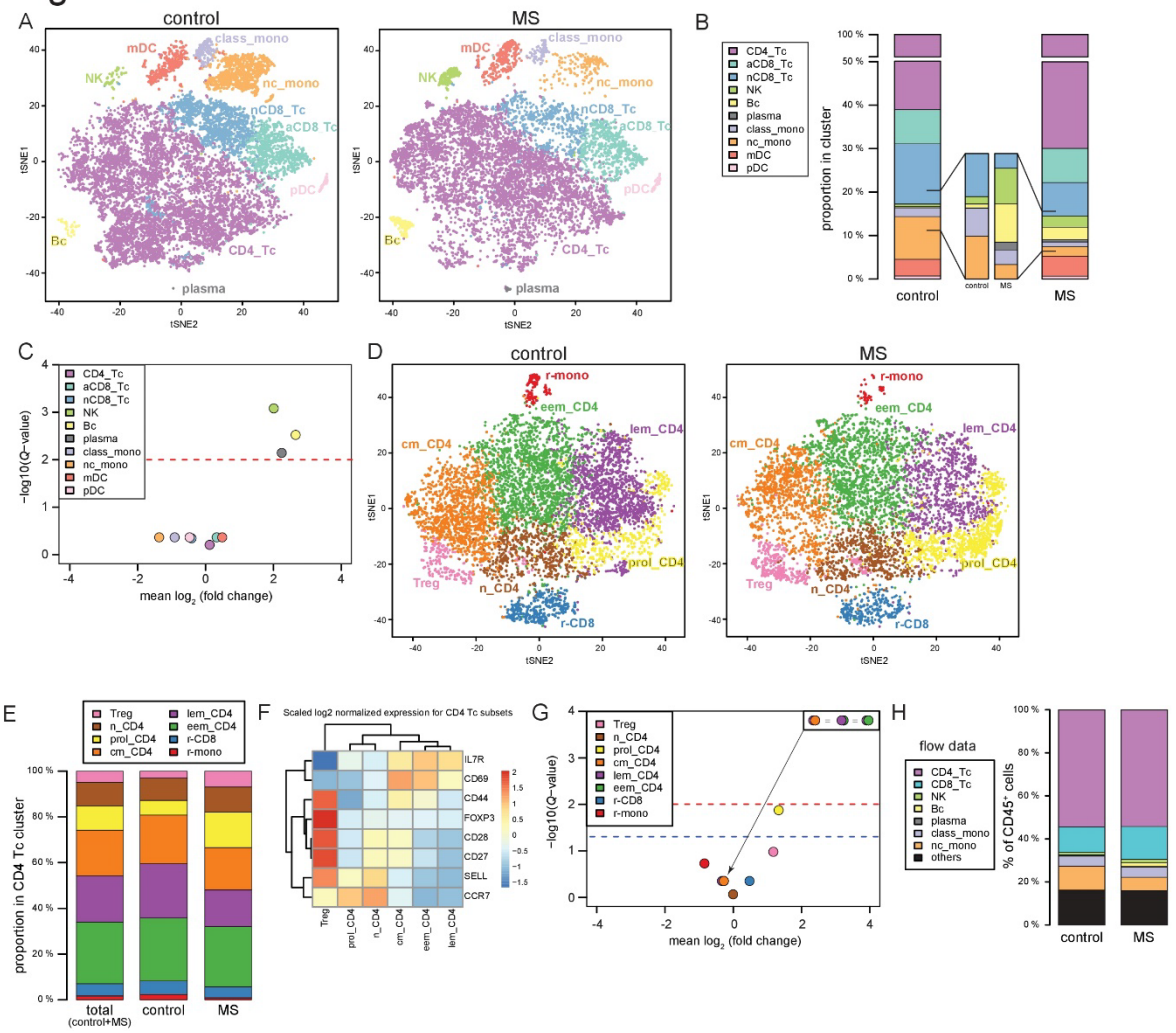


Fig. 2. Unbiased transcriptomics detects NK and B lineage CSF cell expansion in MS.

(A) Condition-specific selections of the t-SNE plot in Figure 1, panel B. The distribution of cell types identified by scRNA-seq in control- (n = 4 donors, 12,705 cells, left plot) and MS-derived (n = 4 donors, 9,652 cells, right plot) CSF cells. (B) Barplots depicting the average proportion of cells in each cluster

in control and MS samples (note split y-axis). Insets highlight abundance of rarer cell types. (C) Volcano plot representing the results of statistical testing for differential cluster abundance between MS vs. control donors. $\log_{10}$-transformed moderated *t*-test *Q*-values (Benjamini-Hochberg) from linear effect modelling on $\log_2$-abundance are plotted against estimated mean $\log_2$ fold change. Horizontal line indicates significance threshold, controlling the FDR < 0.01. (D) Cell profiles from the $CD4^+$ T-cell (*CD4_Tc*) cluster depicted in Figure 1B were re-normalized together and subclustered. A new t-SNE computed for all CD4_Tc cells was subselected to plot 8 subclusters identified across control- (n = 4 donors, 7,764 cells, left plot) and MS-derived (n = 4 donors, 6,749 cells, right plot) CSF samples. (E) Average proportion of cells in each CD4_Tc subcluster (including remainder (r-)CD8 and monocytes (*r-mono*)). (F) Heatmap representing the mean normalized $\log_2$ UMI counts for marker genes (rows) in CD4_Tc subclusters (column) of the dataset depicted in panel D. Rows are Z-normalized so that all marker genes are represented using a common scale. Rows and columns are hierarchically clustered. (G) Volcano plot as in panel C of CD4_Tc subcluster abundance differences between MS vs. control donors. Horizontal lines indicate significance thresholds (blue $Q < 0.05$, red $Q < 0.01$) and inset highlights overlapping symbols representing memory cell types. (H) Barplots representing the proportion of CSF leukocyte subsets identified by flow cytometry in control vs. MS. Cluster key: *CD4_Tc* $CD4^+$ T cells, *aCD8_Tc* / *nCD8_Tc* activated / naïve $CD8^+$ T cells, *NK* natural killer cells, *Bc* B cells, *plasma* plasma cells, *class_mono* / *nc_mono* classical / non-classical monocytes, *mDC* / *pDC* myeloid / plasmacytoid dendritic cells, *Treg* regulatory T helper cells, *n_CD4* naïve, *prol_CD4* proliferating, *cm_CD4* central memory, *lem_CD4* late effector memory, and *eem_CD4* early effector memory $CD4^+$ T cells, *r-CD8* remaining $CD8^+$ T cells, *r-mono* remaining monocytes.

Figure 3

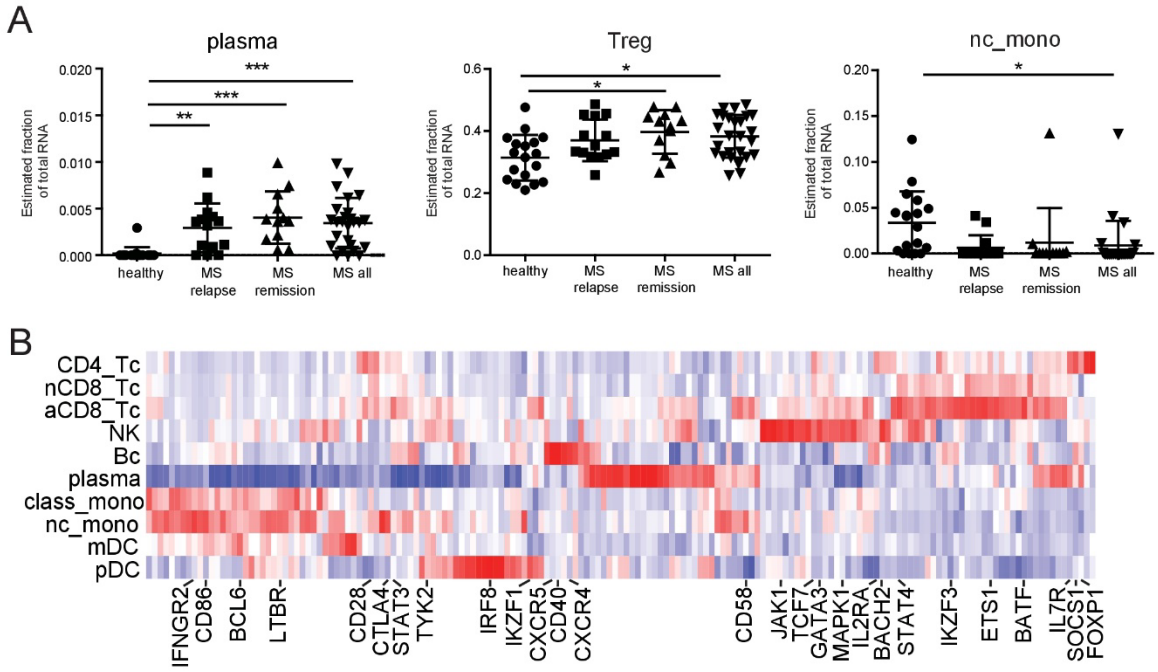


Fig. 3. Deconvolution and interpretation of CSF cell transcriptomes and MS genetics.

(A) Published microarray data of unsorted CSF cells from controls and MS patients in relapse or remission *(19)* were retrieved and cell type deconvolution was performed using cluster-specific gene expression (mean UMI counts) determined by scRNA-seq. Significance was tested applying one-way ANOVA with Tukey's honestly significant differences. (B) Heatmap plotting expression (mean UMI counts) of 167 published MS risk genes (columns) *(40)* against CSF cell cluster (rows). Columns were hierarchically clustered using One minus Pearson correlation and selected gene names are indicated. Cluster names corresponding to Figures 1 and 2 are indicated above each plot. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Figure 4

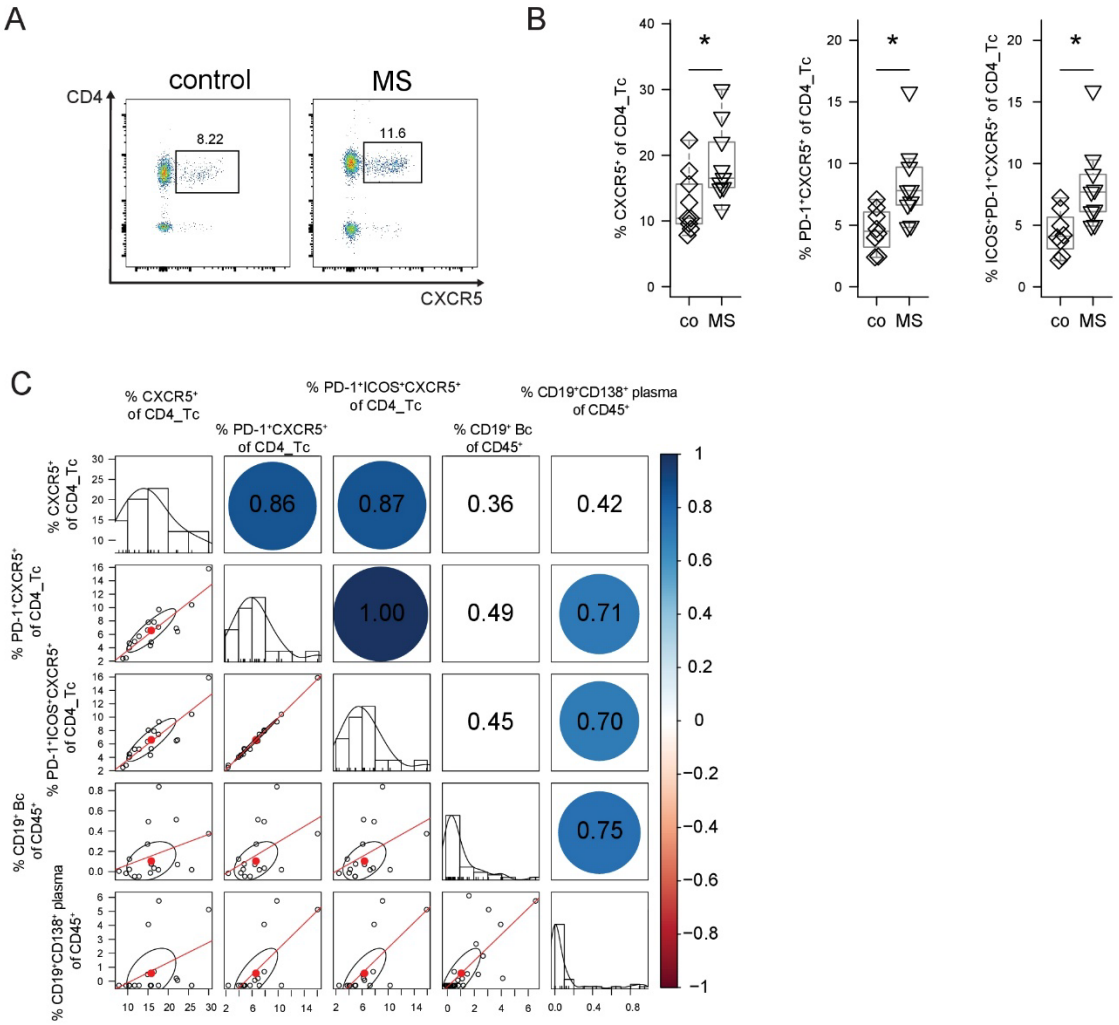


Fig. 4. Increased T follicular helper (TFH) cells in the CSF of MS patients.

(A) Representative flow cytometry dot plot of CSF cells from a control and MS patient stained for CD4 and CXCR5 after gating on live $CD3^+$ cells. (B) The proportion of $CXCR5^+$ (left), of $PD\text{-}1^+CXCR5^+$ (middle), and of $ICOS^+PD\text{-}1^+CXCR5^+$ (right) cells among live $CD3^+CD4^+$ T cells in CSF cells of control (co; n = 9) and MS (n = 9) patients was quantified by flow cytometry. (C) Correlation matrix of $CXCR5^+$ populations and $CD19^+$ and $CD138^+CD19^+$ B lineage cells in the CSF. Blue circles indicate significance and the Pearson correlation coefficient is indicated. * $P < 0.05$

Figure 5

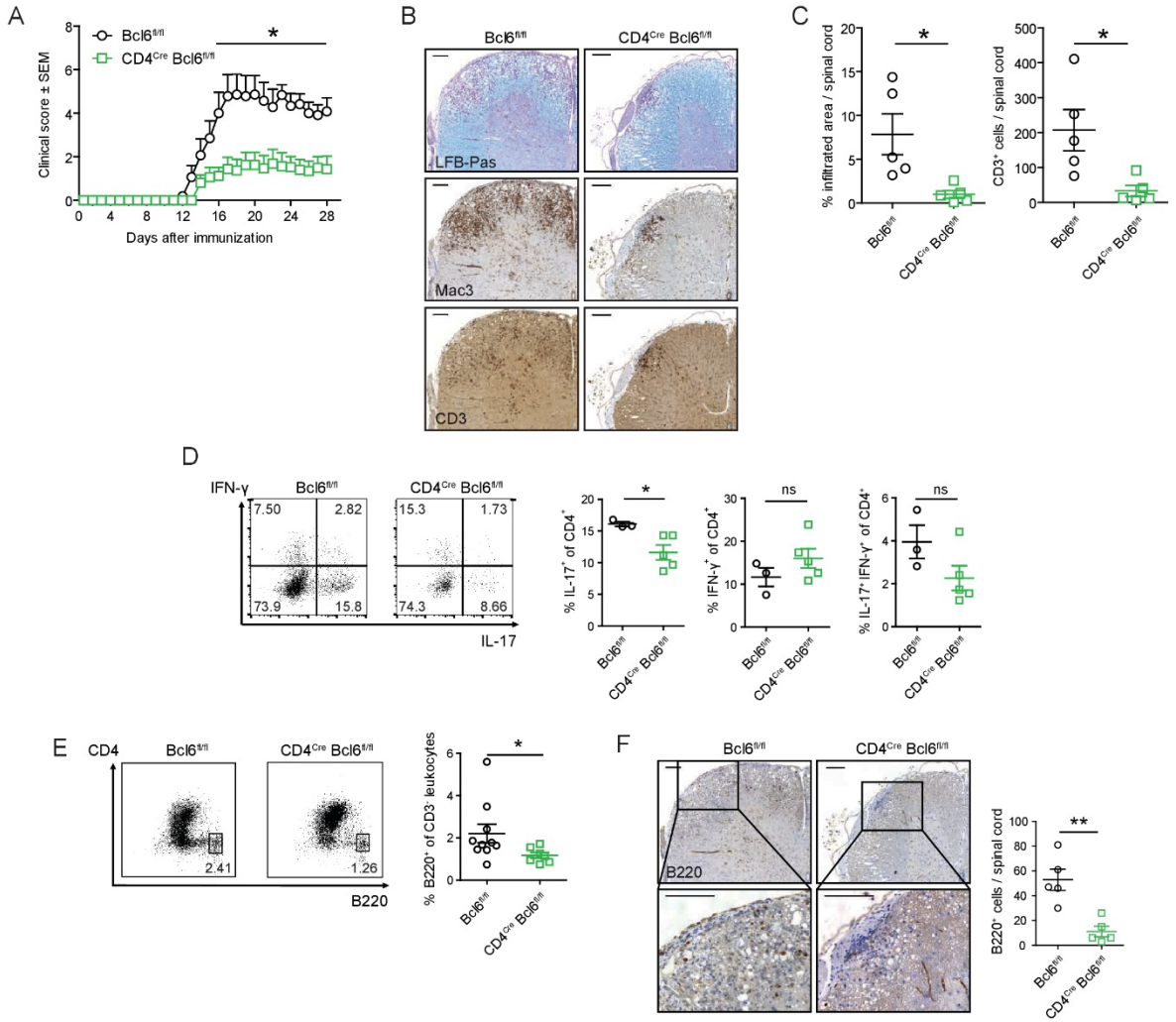


Fig. 5. TFH cells promote neuroinflammation in vivo by expanding B cells.

(A) Active EAE was induced in control Bcl6$^{fl/fl}$ (n = 6) and TFH-deficient CD4^{Cre}Bcl6$^{fl/fl}$ (n = 7) mice using MOG_{35-55} peptide (Methods). Mice were monitored daily for clinical EAE signs. One representative of four independent experiments is shown. (B) At day 28 after EAE induction, spinal cord paraffin cross-sections were stained for LFB-Pas, Mac3 and CD3. (C) The infiltrated area (left) and number of CD3$^+$ cells (right) per spinal cord was quantified manually in a blinded fashion. (D) CNS infiltrating lymphocytes were extracted at peak of EAE and stained for intracellular cytokines (IL-17 and IFN-γ). The proportion of cytokine producing live CD4$^+$ T cells was quantified. (E) CNS infiltrating leukocytes were co-stained for B220 (left) and the proportion of CD3$^-$ B220$^+$ leukocytes was quantified (right). (F) Cross-sections of paraffin embedded spinal cords were stained for B220 (left) and the

proportion of $B220^+$ cells was quantified (right). Scale bars represent 100 μm in panels B and F. * $P <$ 0.05, ** $P < 0.01$, ns not significant.

Supplementary Materials

Supplementary Figure 1

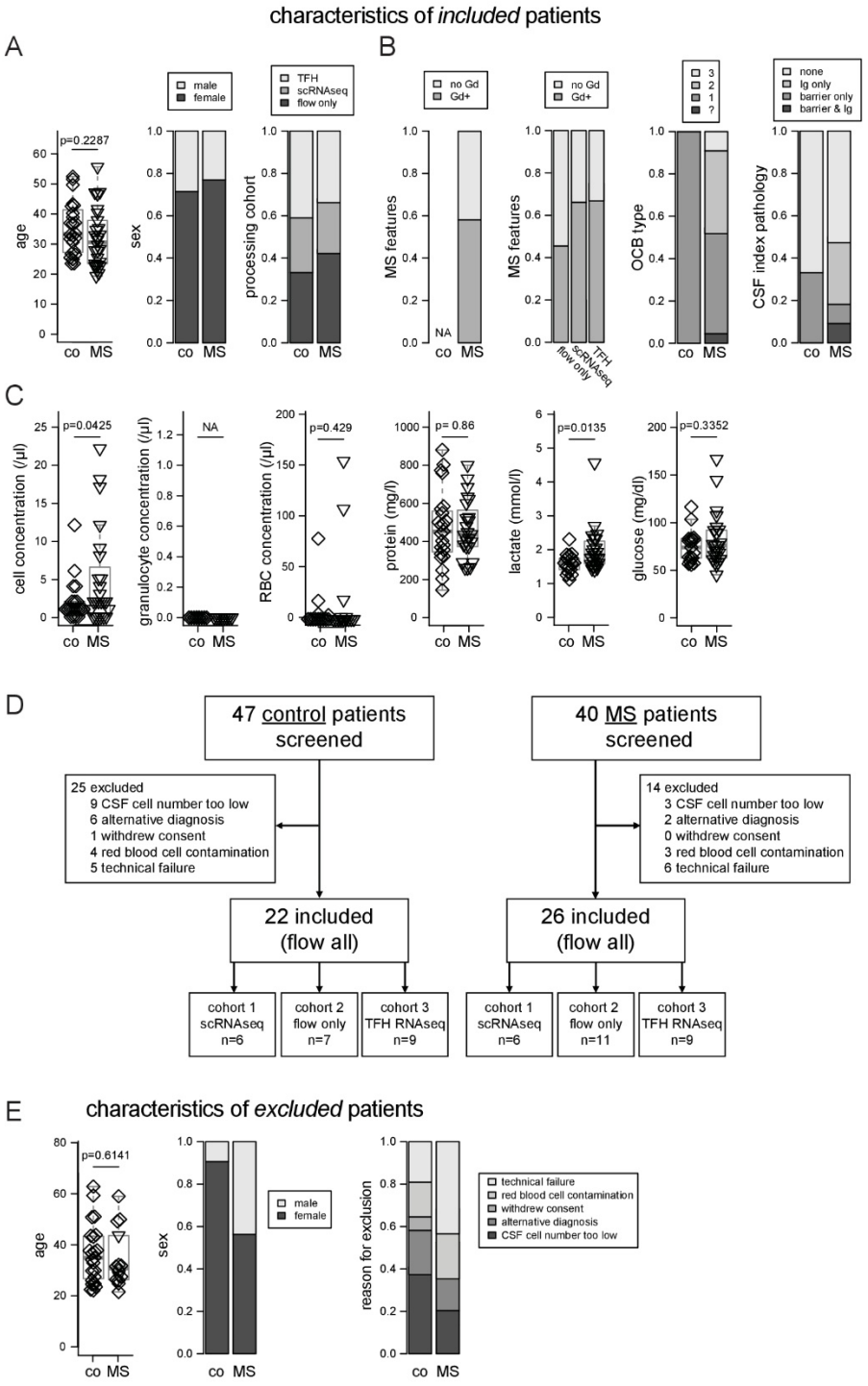


Fig. S1. Patient characteristics.

(A) Clinical characteristics (age, sex) of all control (co, n = 22) and multiple sclerosis (MS, n = 26) patients included into the study after screening are depicted incl.recruitment into the study cohorts. (B) MS patients were classified to either have (Gd+) or not have (no Gd) contrast enhancing lesions in brain or spinal cord detected by magnetic resonance imaging. Oligoclonal bands (OCB) in CSF were classified as being either undetectable (type 1), restricted to CSF (type 2), detected in serum and additionally in CSF (type 3), or not determined (?). CSF/serum indices for albumin and immunoglobulin G (IgG) were calculated. The CSF barrier function (CSF index pathology) was evaluated as being either unaffected (none), showing intrathecal IgG synthesis (Ig only), showing barrier dysfunction (barrier only), or showing both intrathecal IgG synthesis and barrier dysfunction (barrier & Ig). (C) Standard CSF parameters of all study patients including CSF concentrations of total cells, granulocytes, red blood cells (RBC), protein, lactate, and glucose. (D) The study recruitment scheme is depicted. 53% of control and 35% of MS samples were excluded after screening for the reasons indicated. Samples from all patients were divided into three cohorts and all samples were analysed by flow cytometry. Samples in were processed for scRNAseq in cohort 1 and for bulk RNA-seq of sorted T follicular helper cells (TFH) in cohort 3. Samples in cohort 2 were only analysed by flow cytometry. (E) Clinical characteristics (age, sex) of patients excluded after screening and reasons for exclusion are shown. NA not applicable.

Supplementary Figure 2

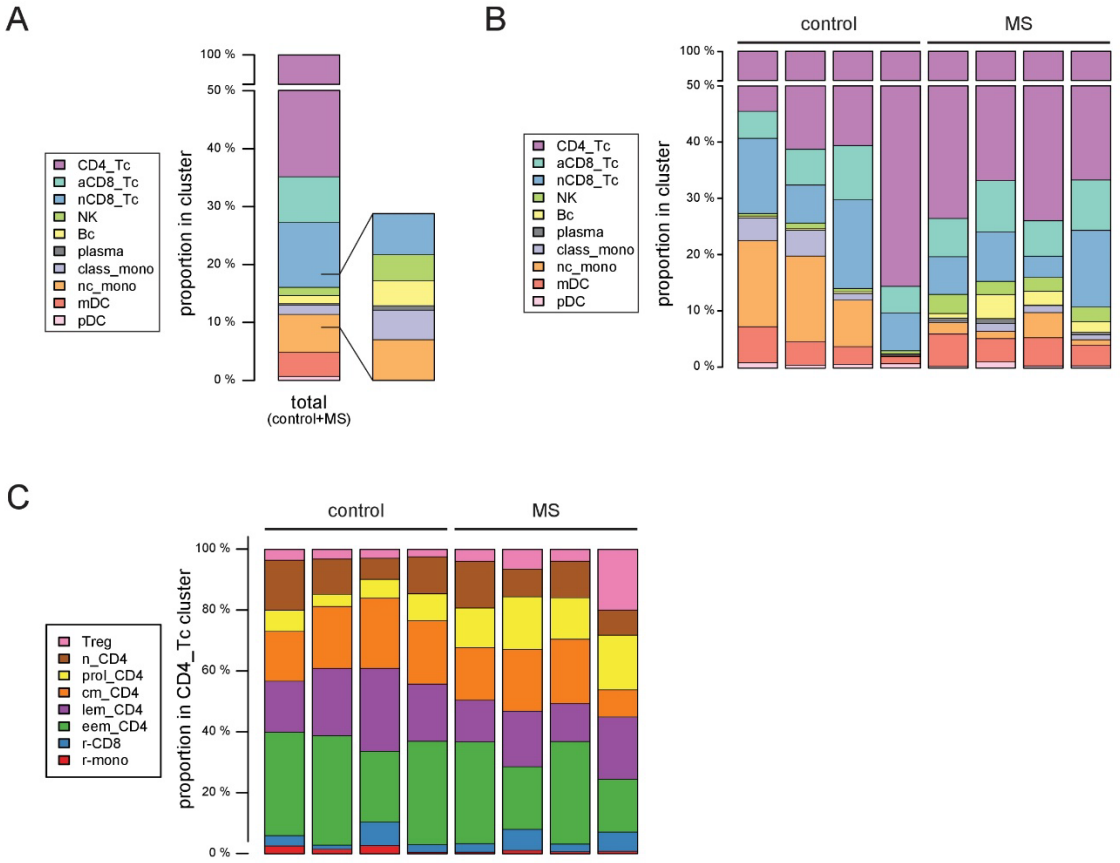


Fig. S2. Average and individual proportions of cell clusters of scRNA-seq samples.

(A) Barplot showing the average proportion of cells in each cluster in all samples (MS and control merged) (note split y-axis). Insets highlight abundance of rarer cell types. (B) Donor-specific proportions of cells in each cluster identified by scRNA-seq (note split y-axis) and (C) in each $CD4^+$ T-cell (*CD4_Tc*) subcluster, for all control and MS patients individually. Cluster key: *CD4_Tc* $CD4^+$ T cells, *aCD8_Tc* / *nCD8_Tc* activated / naïve $CD8^+$ T cells, *NK* natural killer cells, *Bc* B cells, *plasma* plasma cells, *class_mono* / *nc_mono* classical / non-classical monocytes, *mDC* / *pDC* myeloid / plasmacytoid dendritic cells; *Treg* regulatory T helper cells, *n_CD4* naïve, *prol_CD4* proliferating, *cm_CD4* central memory, *lem_CD4* late effector memory, *eem_CD4* early effector memory, *r-CD8* remaining $CD8^+$ T cells, *r-mono* remaining monocytes.

Supplementary Figure 3

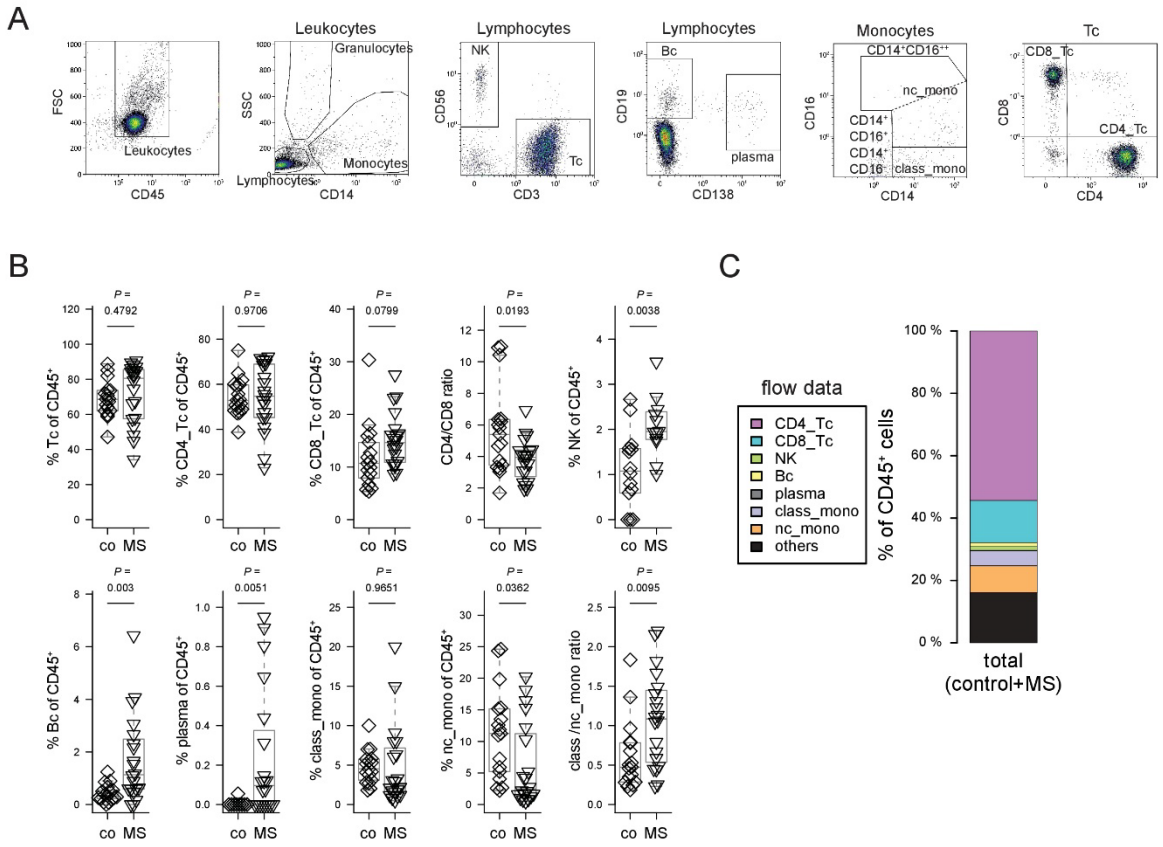


Fig. S3. Flow cytometry characterization of all CSF cell samples.

(A) Representative gating strategy for identifying and quantifying cell types by flow cytometry in the CSF. Population names are indicated next to the respective gates. The proportion of $CD14^{+}CD16^{++}$ monocyte cells in the CSF was very low in accordance with a previous study *(33)*. We therefore merged the $CD14^{+}CD16^{++}$ with the $CD14^{+}CD16^{+}$ cells and named this population as non-classical monocytes (*nc_mono*) for consistency with scRNA-seq naming. (B) Quantification of the indicated cell types in CSF in control (co) and MS patients. All percentages are expressed as proportion of $CD45^{+}$ cells. Samples with less than 500 total $CD45^{+}$ events analysed by flow cytometry were excluded from NK quantification but not from other cell types. Two sided Student's t-test for unrelated samples was used to calculate significance. Exact *P*-values are indicated in the plot. (C) Average proportion of cells in each cluster measured by flow cytometry. Please note the split y-axis and higher magnification inset. *CD4_Tc* $CD4^{+}$ T cells, *CD8_Tc* $CD8^{+}$ T cells, *NK* natural killer cells, *Bc* B cells, *plasma* plasma cells, *class_mono* / *nc_mono* classical / non-classical monocytes.

Supplementary Figure 4

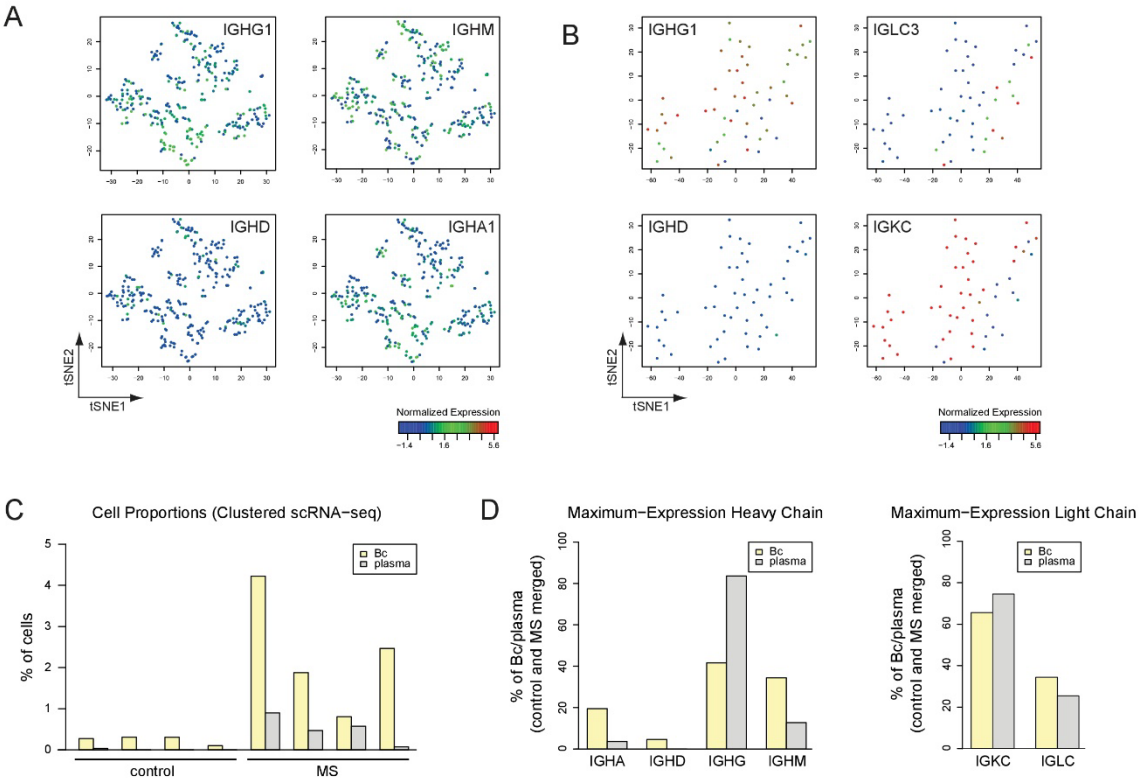


Fig. S4. Late B lineage cells accumulate in the CSF in MS.

(A) Feature plot highlighting the expression level of different heavy chain transcripts in the B cell (Bc) cluster identified in Figure 1B. (B) Feature plot as in panel A showing expression of selected heavy and light chain transcripts in the plasma cell cluster. (C) Proportions for cells in each donor within the B cell or plasma cell cluster. (D) Proportion of Bc and plasma cells expressing indicated heavy (left panel) and light (right panel) chain transcript classes at maximum level (per cell).

Supplementary Figure 5

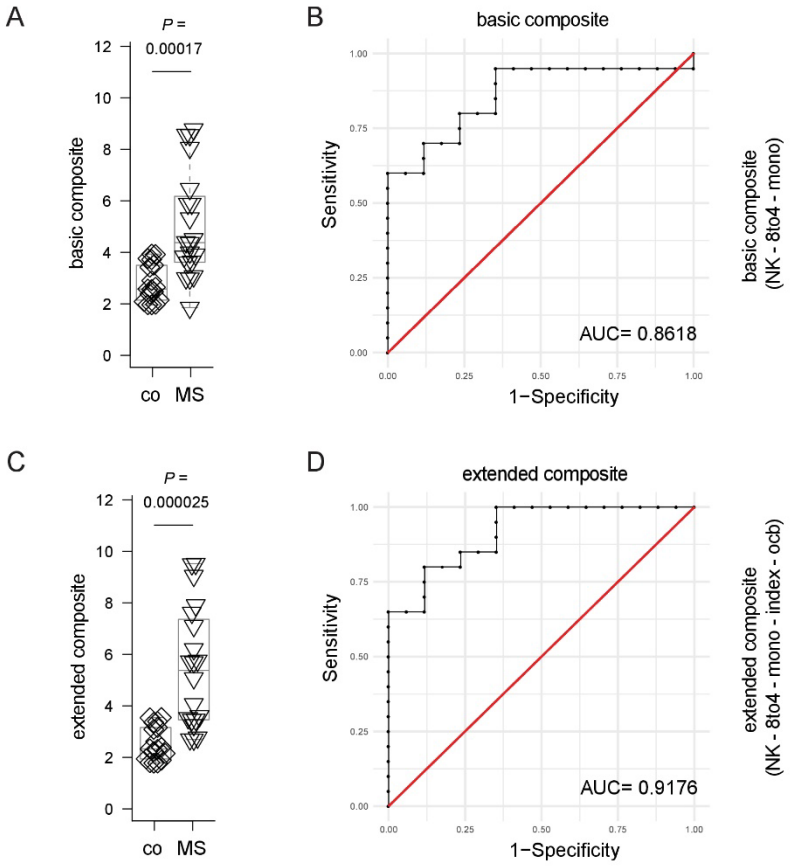


Fig. S5. Evaluating a composite score for diagnosing MS by CSF analysis.

(A) For each samples depicted in the merged flow cytometry data in Suppl. Fig. 3, we calculated a normalized ratio of the proportion of NK to $CD4^+$ T cells, of Bc to $CD4^+$ T cells, of $CD8^+$ to $CD4^+$ T cells, and of $CD14^+CD16^-$ to $CD14^+CD16^+$ monocytes. These four normalized ratios were added to a basic composite score that is depicted in control (co) vs. MS samples. (B) Receiver operator curve (ROC) analysis plotting sensitivity against 1-specificity and the area under the curve (AUC) of the composite. (C) Values of 1 were added to the basic composite depicted in panel A if an elevated immunoglobulin index or oligoclonal bands were detected in the sample. This extended composite score values are depicted by disease status. (D) Receiver operator curve (ROC) analysis of extended composite score.

Supplementary Figure 6

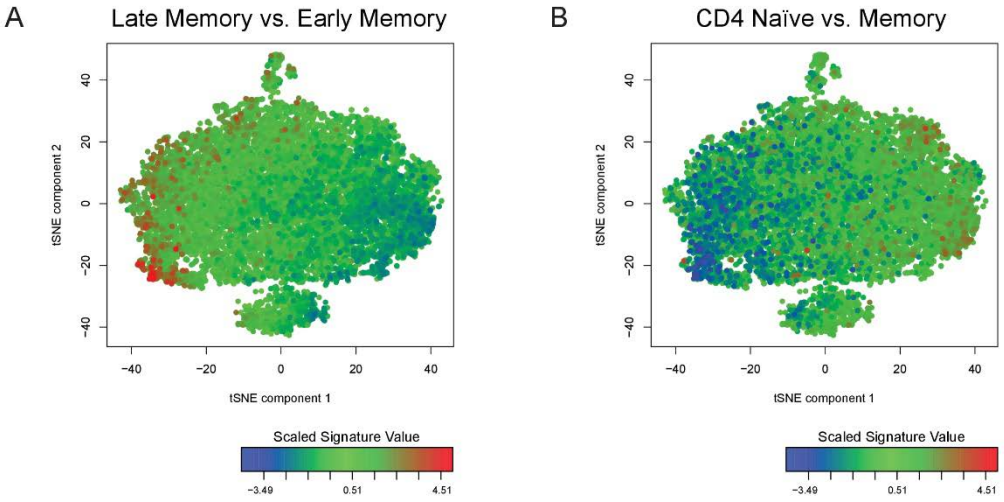


Fig. S6. CD4$^+$ T cells are transcriptionally defined by a continuum.

t-SNE feature plots for CD4$^+$ T cells subclusters representing VISION signatures with significant VISION consistency scores ($P < 0.01$): (A) Late v. early memory signature score from a study on CD8$^+$ cells *(36)*. (B) Naïve v. memory T Cell signature score *(30)*.

Supplementary Figure 7

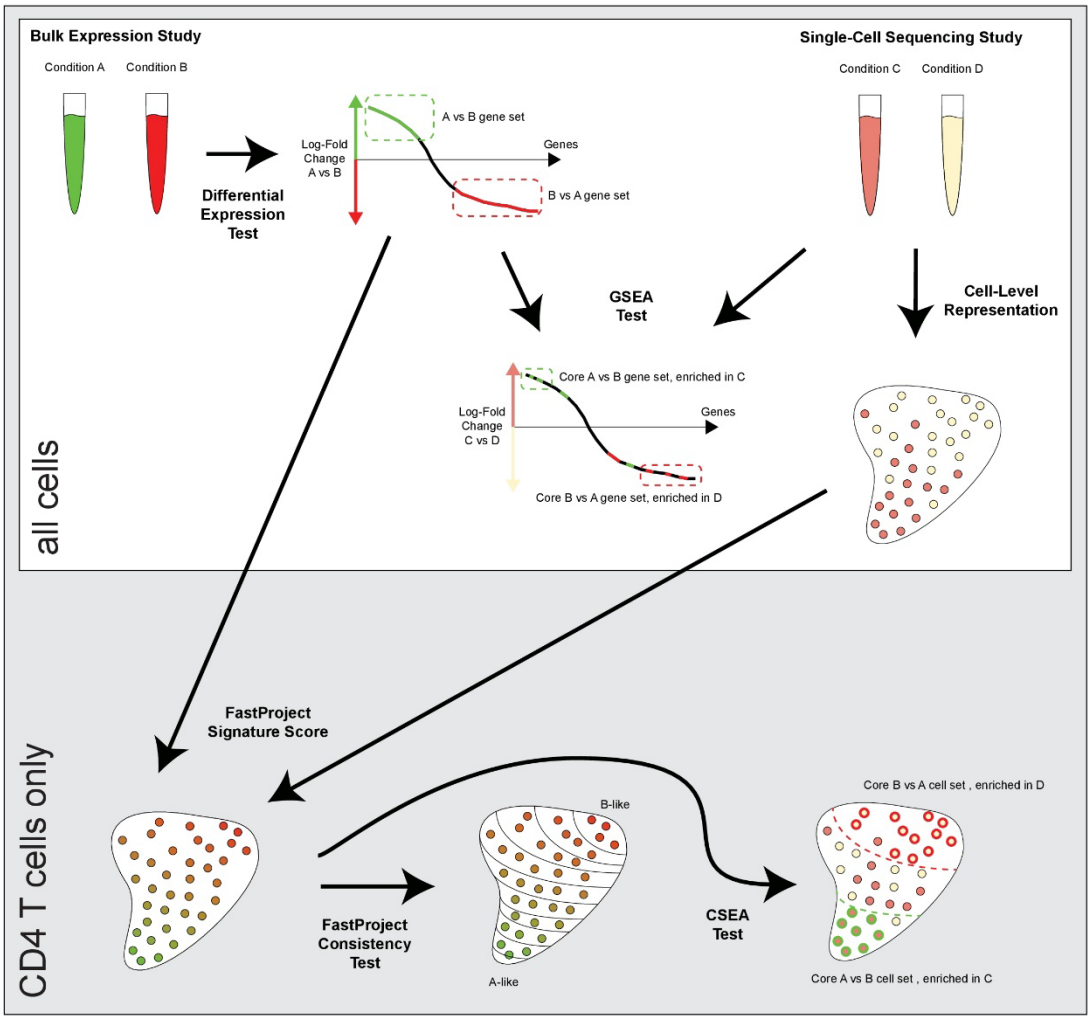


Fig. S7. Scheme of GSEA/VISION/CSEA Analysis.

Publicly available bulk microarray or RNA-seq data are used to identify gene signature sets characterizing immune cell populations. These gene sets are used for either (i) gene set enrichment analysis (GSEA) of our scRNA-seq differential expression results or (ii) single-cell VISION signature scores, input to both VISION Consistency testing and cell set enrichment analysis (CSEA) testing. Further details can be found in the Methods section.

Supplementary Figure 8

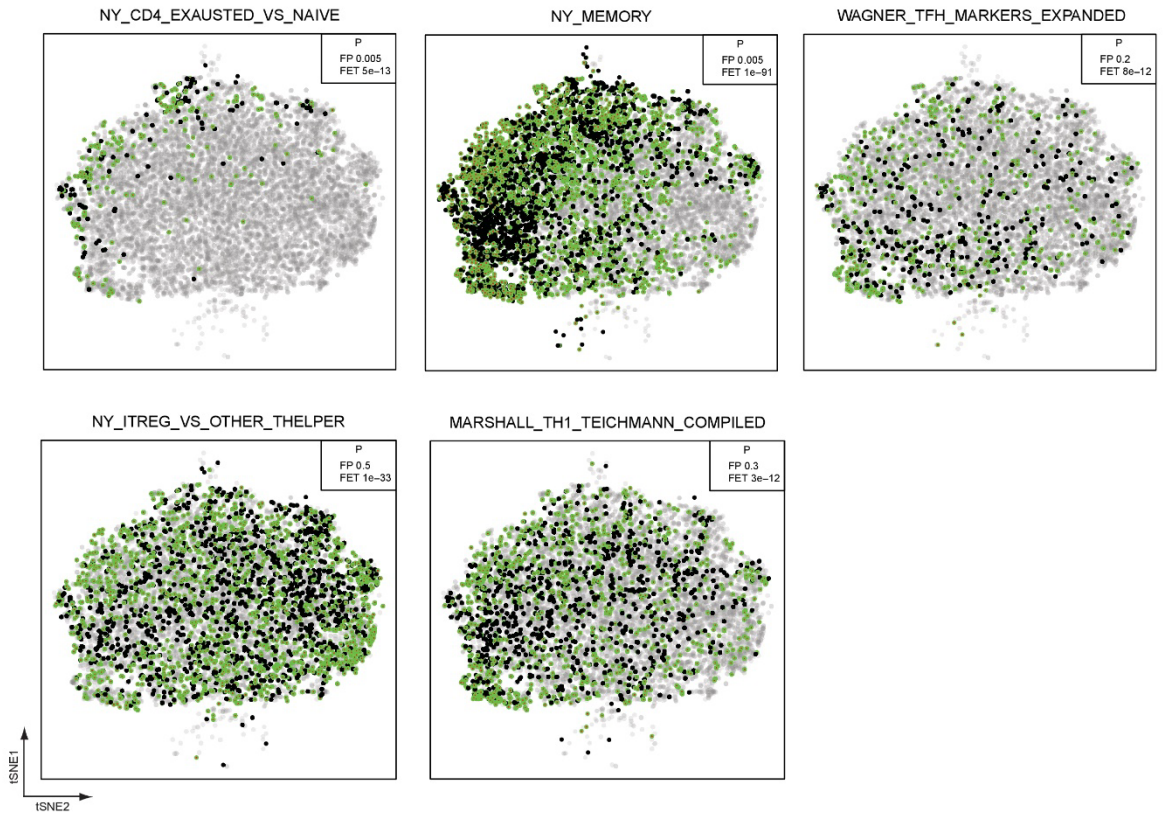


Fig. S8. Cell set enrichment analysis helps identifying disease-specific transcriptional changes.

tSNE plots annotated by representative examples of significant CSEA results for the $CD4^+$ subanalysis. In all cases, MS cells are enriched in the upper tail of the VISION signature distribution. Red points with green outline represent the core MS set driving the signature enrichment, black points are control members of the leading edge cell set. Cells depicted in grey are not members of the leading edge cell set. Exhausted vs. naïve *(44)*, memory *(45)*, T_{FH} *(48, 49)*, iTreg *(47)*, Th1 *(46)*.

Supplementary Figure 9

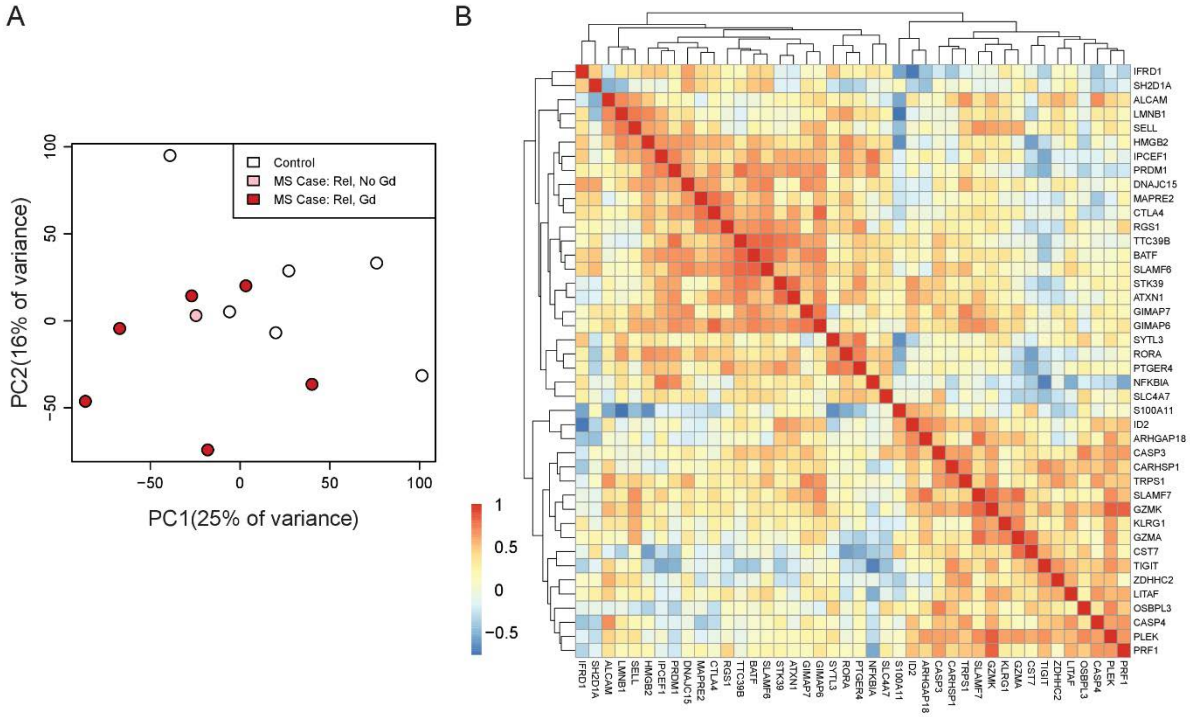


Fig. S9. RNA bulk-seq of TFH cells.

(A) Live $CD3^{+}CD4^{+}CXCR5^{+}$ cells quantified in Fig 5B were flow sorted from the CSF of control (237±107 SD cells) and MS patients (852±691 SD cells), followed by bulk RNA-seq. Scatter plot depicts a principal component analysis (PCA) of normalized TFH gene log-count data, annotated by MS phenotype. (B) Genes differentially expressed in MS vs. controls were sorted by significance score and input into GSEA to identify biologically meaningful enrichments. The most frequent members of core gene sets driving functional enrichments in MS are plotted in a gene-gene Pearson correlation matrix (Methods).

Supplementary Fig 10

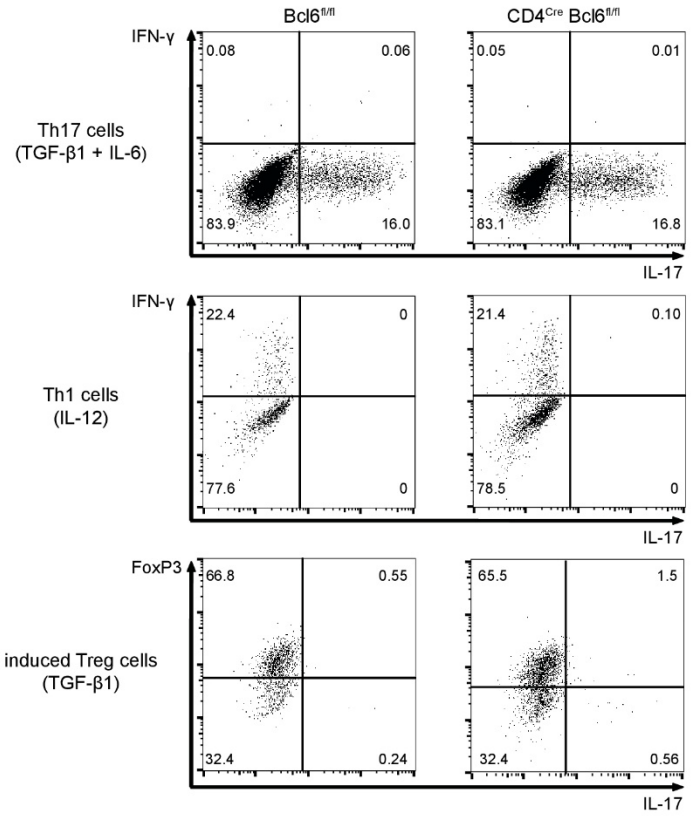
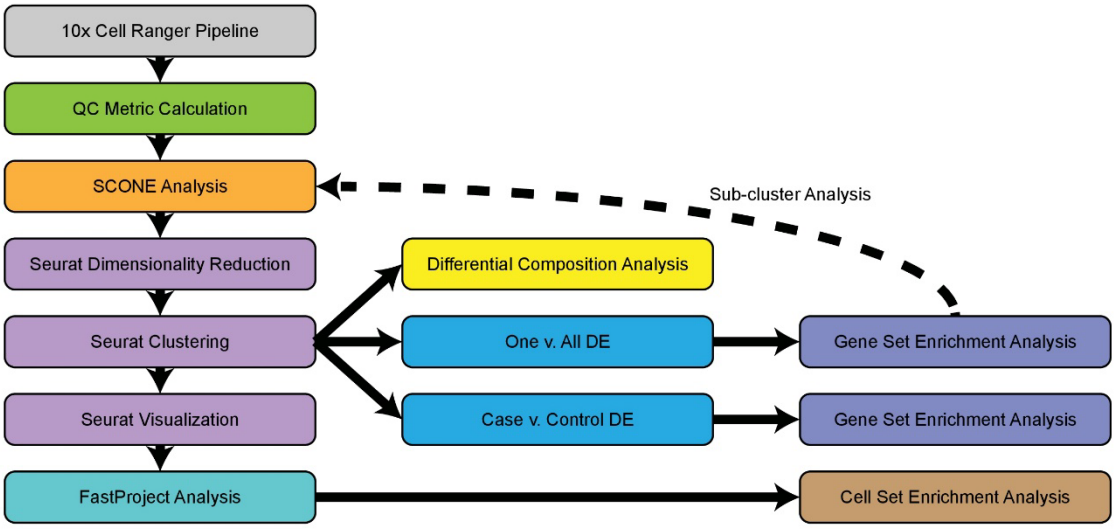


Fig. S10. Bcl6 deficiency does not affect in vitro T helper cell differentiation.

(A) Naïve $CD4^{+}CD62L^{high}CD44^{low}CD25^{-}$ T cells were sorted from $Bcl6^{fl/fl}$ mice and $CD4^{Cre}Bcl6^{fl/fl}$ mice, differentiated in the presence of TGF-β1 and IL-6, or IL-12 alone, or TGF-β1 alone and analysed by intracellular cytokine staining after 4 days in culture.

Supplementary Figure 11



Supplementary Figure 11: Workflow of scRNA-seq analysis.

Scheme depicting the scRNA-seq analysis workflow utilized in this study. Analysis begins with 10X (10X Genomics) Cell Ranger processing and cell-level QC (quality control) metric evaluation, followed by SCONE data filtering and normalization, Seurat dimensionality reduction, clustering and visualization. Results from these analyses are input into VISION for signature calculation and consistency testing. These signatures may be used for CSEA testing. Differential abundance analysis is performed based on the Seurat clustering, and various forms of differential expression testing, including one v. all, “marker” analysis and cluster-specific case v. control analysis are performed using a meta-analysis approach that supports IDR modelling with scRAD tools. GSEA testing is used to ascribe biologic meaning to differential expression results, motivating further subclustering analysis, in which a cluster is analysed using an identical analytical procedure.

Supplementary Table Legends

Table S1. Summarized information about patients in the present study.

Clinical characteristics (average age, sex) of all control (IIH, n=22) and multiple sclerosis (MS, n=26) patients included in the study after screening are depicted. Numbers of excluded patients for each group are also shown. All included patients were divided over three cohorts, cohort 1: CSF samples used for single cell RNA-seq. (6 control vs. 6 MS), cohort 2: CSF samples analysed by flow cytometry only (7 control vs. 11 MS) and cohort 3: CSF samples flow sorted for RNA-seq of $CD3^+CD4^+CXCR5^+$ TFH cells (9 control vs. 9 MS).

Table S2. Standard CSF parameters and MS disease features of patients in the present study.

CSF parameters and MS disease features of all control (IIH, n=22) and multiple sclerosis (MS, n=26) patients included in the study are depicted. All included patients were divided over three cohorts, cohort 1: CSF samples used for single cell RNA-seq. (6 control vs. 6 MS), cohort 2: CSF samples analysed by flow cytometry only (7 control vs. 11 MS) and cohort 3: CSF samples flow sorted for RNA-seq of $CD3^+CD4^+CXCR5^+$ TFH cells (9 control vs. 9 MS). All MS patients were classified if they had a relapse at CSF collection, if they had (Gd+) or not had (no Gd) contrast enhancing lesions in brain or spinal cord or if they had other MS typical characteristics observed by magnetic resonance imaging (MRI). Oligoclonal bands (OCB) in CSF were classified as being either undetectable (type 1), or restricted to CSF (type 2), or detected in serum and additionally in CSF (type 3), or not determined (?). The CSF barrier function was evaluated as being either unaffected (none), or showing intrathecal IgG synthesis (Igonly), or showing barrier dysfunction (barrier only), or showing both intrathecal IgG synthesis and barrier dysfunction (barrier & Ig) or being unclassified (unknown). Standard CSF parameters of all study patients including CSF concentrations of protein, lactate, glucose, total cells, granulocytes and red blood cells (RBC) are also depicted.

Table S3. Technical information of scRNA-seq results.

Technical information on scRNA-seq results of all patients (Control, n=4 and MS, n=4) included in the study are depicted. Depicted is the number of samples used for scRNA-seq (number of samples for 10x), the total number of expected cells based on counting cells included in each sample multiplied by the approximate capture rate of the 10x system (total number of expected cells), the total number of measured cells after sequencing and genome alignment (total number of measured cells), average number of measured cells per sample (average number of measured cells), average number of detected reads per cell (reads per cell) and average number of detected genes per cell (genes per cell) used for downstream analysis. The total and average number of cells measured within the $CD4^+$ T cell (CD4_Tc) cluster is also depicted.

Table S4. Merged results of the scRNA-seq analysis.

Genes most differentially expressed in clusters identified in the first clustering including all cells (All) and in the secondary $CD4^+$ T cell clustering (CD4) are listed. Depicted are the Ensembl IDs of the Genes tested (Ensembl ID), the common Gene names (Gene Symbol), the cluster analysed (Cluster), the median log2 fold change (Median Log2 Fold Change), the irreproducible discovery rate (IDR), the statistical significance (Meta-analysis *P*-value) and the false discovery rate (Meta-analysis FDR). Candidate genes were defined as reaching a threshold of either log2 fold change and IDR and FDR, or IDR and FDR, or Median Log2 Fold Change and FDR. Additionally, candidate genes from different DE analysis comparing one cluster to all others (One vs. All (Marker)) and comparing same clusters between Multiple sclerosis patients (MS) and Control patients (IIH) (MS vs. IIH (Exposure)) are depicted.

Table S5. Gene set enrichment analysis (GSEA) results for genes differentially expressed by clusters.

Gene set enrichment analysis (GSEA) was performed on all marker genes for every cluster after the first clustering using all cells. Depicted are enriched Gene Sets (Signature), reference links to the GSEA data base (Origin), GSEA Enrichment Scores (EScore), statistical significance (PValue), simulated *P*-values using bonferroni correction (sim_p_bonferroni), Cluster analysed (Cluster), differential analysis type (DEType), direction of Gene set enrichment (Sign) and signature containing collections

(signatures_NY_private, c1.all.v6.1.symbols, c2.all.v6.1.symbols, c2.cgp.v6.1.symbols, c2.cp.biocarta.v6.1.symbols, c2.cp.kegg.v6.1.symbols, c2.cp.reactome.v6.1.symbols, c2.cp.v6.1.symbols, c3.all.v6.1.symbols, c3.mir.v6.1.symbols, c3.tft.v6.1.symbols, c4.all.v6.1.symbols, c4.cgn.v6.1.symbols, c4.cm.v6.1.symbols, c5.all.v6.1.symbols, c5.bp.v6.1.symbols, c5.cc.v6.1.symbols, c5.mf.v6.1.symbols, c6.all.v6.1.symbols, c7.all.v6.1.symbols, h.all.v6.1.symbols, msigdb.v6.1.symbols).

Table S6. Gene set enrichment analysis (GSEA) results for genes differentially expressed in MS vs. control samples.

Gene set enrichment analysis (GSEA) was performed on all differentially expressed (MS vs. control) marker genes for every cluster after the first clustering (All). Depicted are enriched Gene Sets (Signature), reference links to the GSEA data base (Origin), GSEA Enrichment Scores (EScore), statistical significance (*P*-value), simulated *P*-values using bonferroni correction (sim_p_bonferroni), Cluster analysed (Cluster), differential analysis type (DEType), direction of Gene set enrichment (Sign) and signature containing collections (signatures_NY_private, c1.all.v6.1.symbols, c2.all.v6.1.symbols, c2.cgp.v6.1.symbols, c2.cp.biocarta.v6.1.symbols, c2.cp.kegg.v6.1.symbols, c2.cp.reactome.v6.1.symbols, c2.cp.v6.1.symbols, c3.all.v6.1.symbols, c3.mir.v6.1.symbols, c3.tft.v6.1.symbols, c4.all.v6.1.symbols, c4.cgn.v6.1.symbols, c4.cm.v6.1.symbols, c5.all.v6.1.symbols, c5.bp.v6.1.symbols, c5.cc.v6.1.symbols, c5.mf.v6.1.symbols, c6.all.v6.1.symbols, c7.all.v6.1.symbols, h.all.v6.1.symbols, msigdb.v6.1.symbols).

Table S7. VISION and Cell set enrichment analysis (CSEA) results for T cell signatures.

Cell set enrichment analysis (CSEA) was performed on all $CD4^+$ T cells after removing residual clusters. Columns in this sheet include i) signature set (Signature), ii) VISION Z-score (VISIONZ), iii) Benjamini-Hochberg *Q*-values from permutation-based VISION *P*-values (VISIONQ), iv) positive signature MS enrichment score (csea_sign1_MS_enriched_e_score), v) simulated *P*-values for the positive MS enrichment score, adjusted using the Bonferroni correction (csea_sign1_MS_enriched_sim_p_bonferroni), vi) number of cells in the positive leading edge for

enrichment in MS (csea_sign1_MS_enriched_leading_edge_size), vii-ix) analogous columns for negative signature MS enrichment (csea_sign-1_MS), x-xii) analogous columns for positive signature control (IIH) enrichment (csea_sign1_IIH), and xiii-xv) analogous columns for negative signature control (IIH) enrichment (csea_sign-1_IIH).

Table S8. Flow sorting related information.

Statistics of follicular T helper (TFH) cells analysed and sorted out of CSF using fluorescence activated cell sorting (FACS) for all control (n=9) and multiple sclerosis (MS, n=9) patients. Depicted are the number of processed samples (# samples processed), the number of CSF TFHs analysed (CSF TFH (#)) and the average number of TFHs per sample (average ± SD).

Table S9. Differentially expressed genes and gene set enrichment analysis (GSEA) in CSF-derived TFH cells in MS vs. control patients.

Per-gene differential expression (DE) analysis was performed on TFH cells sorted out of the CSF from MS and control patients. The limma::topTable results for disease effect estimation are tabulated in the "DE" sheet. Columns in this sheet include i) the symbol for the gene tested (Gene Symbol), ii) $\log_e$ fold-changes (MS vs. control) in normalized expression (logFC), iii) average normalized $\log_e$-expression (AveExpr), iv) moderated t-values (t), v) statistical significance (*P*-Value), vi) Benjamini-Hochberg *Q*-value and vii) log-odds that the gene is differentially expressed (B). The B-value is the log-odds that the gene is differentially expressed.

Gene set enrichment analysis (GSEA) was performed on significance scores derived from comparisons of TFH cells sorted out of the CSF from MS and Control patients; results are shown in the "GSEA" sheet. Columns in this sheet include i) the gene set origin, e.g. an experimental comparison from which the gene set is derived (Signature), signature subset, indicating whether the subset is up-regulated or down-regulated in the external comparison (Signature subset), gene set enrichment score for genes with high significance scores (EScore), simulated *P*-values, adjusted using the Bonferroni correction for multiple testing (sim_p_bonferroni), names of genes driving the gene set enrichment (Core Genes), number of genes driving the gene set enrichment (Number of Core Genes).

Table S10. Deconvolution results.

scRNA-sequencing data were used to deconvolute the cell composition of already published bulk sequencing data. Depicted are input samples deconvolution was performed on (Input Sample) and the percentile composition of the different populations identified by scRNA-seq. Cluster key: a*CD8_Tc / nCD8_Tc* activated / naïve CD8 T cells, *Bc* B cells, *class_mono* classical monocytes, *mDC* myeloid dendritic cells,, *nc_mono* non-classical monocytes, *NK* Natural killer cells,, *pDC* plasmacytoid dendritic cells, *plasma* plasma cells, *Treg* regulatory T cells ; statistical significance (*P*-value), statistical correlation (Pearson Correlation) and Root Mean Square Error (RMSE).