

Twin study reveals non-heritable immune perturbations in multiple sclerosis

<https://doi.org/10.1038/s41586-022-04419-4>

Received: 18 March 2021

Accepted: 4 January 2022

Published online: 16 February 2022

Open access

 Check for updates

Florian Ingelfinger^{1,2,13}, Lisa Ann Gerdes^{3,4,5,13}, Vladyslav Kavaka^{3,4}, Sinduya Krishnarajah¹, Ekaterina Friebe¹, Edoardo Galli^{1,12}, Pascale Zwicky¹, Reinhard Furrer^{6,7}, Christian Peukert⁸, Charles-Antoine Dutre^{9,10}, Klara Magdalena Eglseer^{3,4}, Florent Ginhoux¹¹, Andrea Flierl-Hecht^{3,4,5}, Tania Kumpfel^{3,4,5}, Donatella De Feo¹, Bettina Schreiner^{1,2}, Sarah Mundt¹, Martin Kerschensteiner^{3,4,5}, Reinhard Hohlfeld^{3,4,5}, Eduardo Beltrán^{3,4,5,14} & Burkhard Becher^{1,14}✉

Multiple sclerosis (MS) is a chronic inflammatory disorder of the central nervous system underpinned by partially understood genetic risk factors and environmental triggers and their undefined interactions^{1,2}. Here we investigated the peripheral immune signatures of 61 monozygotic twin pairs discordant for MS to dissect the influence of genetic predisposition and environmental factors. Using complementary multimodal high-throughput and high-dimensional single-cell technologies in conjunction with data-driven computational tools, we identified an inflammatory shift in a monocyte cluster of twins with MS, coupled with the emergence of a population of IL-2 hyper-responsive transitional naive helper T cells as MS-related immune alterations. By integrating data on the immune profiles of healthy monozygotic and dizygotic twin pairs, we estimated the variance in CD25 expression by helper T cells displaying a naive phenotype to be largely driven by genetic and shared early environmental influences. Nonetheless, the expanding helper T cells of twins with MS, which were also elevated in non-twin patients with MS, emerged independent of the individual genetic makeup. These cells expressed central nervous system-homing receptors, exhibited a dysregulated CD25–IL-2 axis, and their proliferative capacity positively correlated with MS severity. Together, our matched-pair analysis of the extended twin approach allowed us to discern genetically and environmentally determined features of an MS-associated immune signature.

MS is the most common neurological disorder affecting young adults with rising global incidence and prevalence, particularly in Western countries¹. Despite numerous genome-wide association and epidemiological studies^{2–4}, the aetiology of MS remains largely unknown. Yet, environmental cues associated with increased risk of developing MS have been established, and include factors acting prenatally^{5,6}, during adolescence and adulthood (for example, viral infections, low levels of vitamin D, and tobacco smoke)⁷. Conversely, a registry-based study investigated the familial recurrence rate of MS and revealed that monozygotic twin pairs displayed the highest familial risk (17% age-adjusted risk for the unaffected twin) of developing MS, indicating a strong heritable effect⁸. Accordingly, over 200 risk loci with moderate-to-subtle effects have been described, including *HLA-DRB1*15:01*, *IL2RA* and *IL7R* genes^{3,9}. Given that only 18–24% of MS heritability can be explained by known risk loci^{2,10}, it is evident that the aetiology of MS involves a complex interplay between polygenic risk variants and environmental triggers. Hence,

investigating how genetic predisposition and environmental triggers shape the interactions of individual immune cells is vital to understand the pathophysiology of autoimmune diseases including MS.

Previous attempts linking individual risk factors with functional perturbation on a single-cell level have given some insight into the inflammatory processes underlying MS^{8,11–13}. Similarly, several studies have compared the immune profiles of patients with MS to healthy individuals or other control individuals with a disease^{14–16}. Whereas these cross-sectional analyses revealed potential biomarkers, they cannot control for the effects of the genetic heterogeneity underlying MS susceptibility required for a patient to develop MS and its potential influence on immune cell perturbations.

Here we conducted an in-depth pairwise analysis of the systemic immune compartment of 61 monozygotic twin pairs discordant for MS, in which both siblings carry the same genetic and early-life environmental risk for the disease, yet only one is affected by MS. This approach thus eliminated the majority of bias attributed to variable genetic

¹Institute of Experimental Immunology, University of Zurich, Zurich, Switzerland. ²Department of Neurology, University Hospital Zurich, Zurich, Switzerland. ³Institute of Clinical Neuroimmunology, University Hospital, LMU Munich, Munich, Germany. ⁴Biomedical Center (BMC), Faculty of Medicine, LMU Munich, Martinsried, Germany. ⁵Munich Cluster of Systems Neurology (SyNergy), Munich, Germany. ⁶Department of Mathematics, University of Zurich, Zurich, Switzerland. ⁷Department of Computational Science, University of Zurich, Zurich, Switzerland. ⁸Department of Strategy, Globalization and Society, University of Lausanne, Lausanne, Switzerland. ⁹Gustave Roussy Cancer Campus, Villejuif, France. ¹⁰Institut National de la Santé Et de la Recherche Médicale (INSERM) U1015, Equipe Labellisée—Ligue Nationale contre le Cancer, Villejuif, France. ¹¹Singapore Immunology Network, A*STAR, Singapore, Singapore. ¹²Present address: Neurologic Clinic and Polyclinic, University Hospital Basel, Department of Biomedicine, University of Basel, Basel, Switzerland. ¹³These authors contributed equally: Florian Ingelfinger, Lisa Ann Gerdes. ¹⁴These authors jointly supervised this work: Eduardo Beltrán, Burkhard Becher. ✉e-mail: becher@immunology.uzh.ch

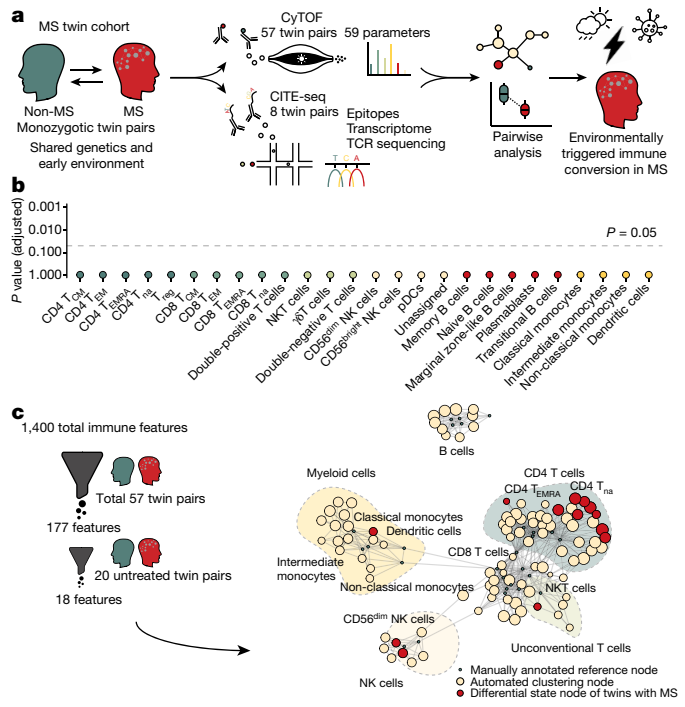


Fig. 1 | Twins discordant for MS exhibit differences in circulating T_H cell and myeloid compartments. **a**, PBMCs from monozygotic twin pairs discordant for MS were analysed by CyTOF (MS; $n = 57$) and CITE-seq ($n = 8 + 2$ healthy twins) to reveal an environmentally triggered immune conversion in MS. **b**, Lollipop plot showing P values for the two-sided paired non-parametric Wilcoxon signed-rank test with a false discovery correction according to the Benjamini-Hochberg approach of immune populations in the reference framework for untreated twins with MS ($n = 20$) and controls ($n = 20$). The dashed line indicates a 5% significance threshold. **c**, Automated analysis using diffcyt was performed by applying two filters screening for immune features that are different between all twin pairs ($n = 57$) and twin pairs in which the twin with MS remained untreated ($n = 20$) (left). Network visualization (right) of unsupervised FlowSOM clusters yielded by diffcyt analysis (beige; automated clustering nodes), the differential state nodes for twins with MS for which features appeared significantly different in both filters (red) and the nodes of the manually annotated reference framework defined in Extended Data Fig. 1c (green) are also shown. The dot size corresponds to the population frequency among total leukocytes. pDCs, plasmacytoid dendritic cells; T_{eff}, effector T cell; T_{na}, naive T cell; T_{reg}, regulatory T cell.

and early environmental influences in a heterogeneous population^{17,18}. We combined the high-throughput of mass cytometry, facilitating the analysis of 57 twin pairs, with cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) of eight selected twin pairs to obtain a comprehensive overview of epitopes, transcriptome and T cell receptor (TCR) clonotypes. Substantial MS-associated alterations were largely restricted to the myeloid and helper T (T_H) cell compartments. Our study thereby resolves how the immune systems of twins, who share the same genetic and early environmental risk factors, can diverge towards distinct clinical phenotypes.

MS signature in T_H and myeloid cells

We began by comparing the peripheral blood immune cell populations of 57 patients with MS and their unaffected monozygotic twin siblings (Extended Data Table 1, Supplementary Tables 1, 2). We applied cytometry by time-of-flight (CyTOF) to samples of peripheral blood mononuclear cells (PBMCs) to define the abundance of different cell types capturing 59 cellular parameters at the single-cell level (Fig. 1a, Supplementary Table 3). By combining data-driven and hypothesis-driven

analyses, we created a cellular reference framework with well-defined canonical immune cell populations.

High-resolution clustering using FlowSOM and manual merging assigned all cells to a canonical immune subset, creating a framework of reference nodes that served as the basic substrate for the data-driven analysis (Extended Data Fig. 1a-c). Next, pairwise comparison of the frequencies of immune cell subsets in twins affected and unaffected by MS was performed, thereby controlling for the effects of genetic and shared early environmental (for example, prenatal and early childhood) influences. Although we uncovered significant differences in the frequency of classical monocytes in twins with MS (Extended Data Fig. 1d), we discovered that this was primarily driven by disease-modifying therapies (Fig. 1b). Thus, after eliminating genetic and early environmental sources of variance, monozygotic twins discordant for MS exhibit comparable frequencies of immune cell subsets in peripheral blood.

For a data-driven agnostic approach, which does not categorize cells into canonical subsets based on previous knowledge, the recently introduced diffcyt toolset based on the widely established Bayes moderated tests utilized in transcriptomic studies was applied. Markers were divided into lineage markers, which had served as the foundation for FlowSOM clustering, and cell-state markers including activation and trafficking molecules and cytokines (Supplementary Table 3). On the basis of this classification, diffcyt generated a list of 1,400 immune features, in which each feature represents the expression of a cell-state marker in a given immune cluster. Those features were then filtered by meeting two conditions: they distinguished twins with MS from their unaffected twin siblings across all twin pairs, and they were not elicited by disease-modifying therapy (Fig. 1c). Eighteen different immune features fulfilled both conditions in twins with MS and corresponding immune clusters were termed ‘differential state nodes’ (Extended Data Table 2). Twelve out of the 18 features localized within the T_H cell region, supporting previous observations^{13,14}, one feature was close to the unconventional T cell reference nodes, three were close to the natural killer (NK) cell reference nodes and the remaining two were in the myeloid cell region (Fig. 1c, right). The most significant immune alterations were observed in the T_H cell and myeloid cell compartment (Extended Data Table 2). The conjunction of data-driven and previous knowledge-driven approaches has thus demonstrated that the clinical manifestation of MS, independent of the genetic predisposition, is predominantly determined in the T_H cell and myeloid compartments.

CCR2-CSF2R are elevated in MS monocytes

Mononuclear phagocytes constitute a dominant fraction of leukocytes infiltrating the central nervous system (CNS) found in lesions of patients with MS^{19,20} and their inhibition has been reported to prevent neuroinflammation in mice²¹⁻²³. Both MS-related features of the myeloid compartment were found in the same differential state node and exhibited higher levels of the receptor for monocyte chemoattractant protein 2 (CCR2) and the granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor-specific subunit (CD116) in twins with MS (Fig. 2a). This differential state phagocyte node in twins with MS phenotypically resembled classical monocytes (Fig. 2a, Extended Data Fig. 2a). However, CD14 expression was dim and CD16 was absent, obstructing their unambiguous assignment to a canonical immune population (Extended Data Fig. 2a). The strong positive correlation between CD116 and CCR2 expression in monocytes within the differential state phagocyte node suggests a common regulation of both molecules (Extended Data Fig. 2b). Elevated expression of CCR2 and CD116 was especially pronounced in untreated twin pairs, suggesting that this particular MS-related perturbation is susceptible to disease-modifying therapies (Fig. 2a, right, Extended Data Fig. 2c).

To fully understand the identity of the cells within the differential state myeloid node, we enriched the cohort of twins with MS with InfinityFlow data of PBMCs for 347 detected surface markers derived from a

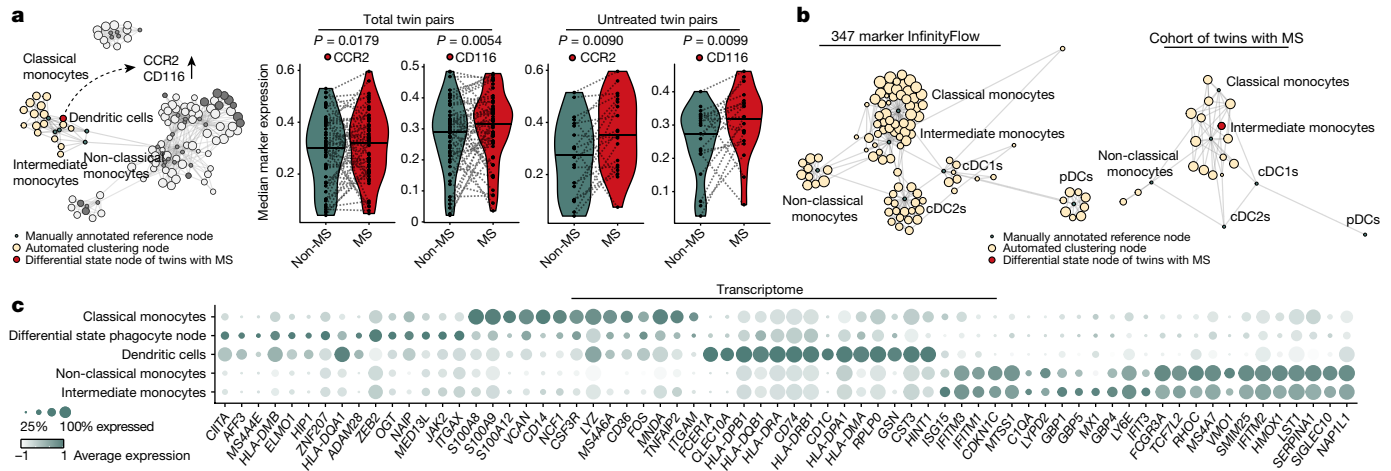


Fig. 2 | The myeloid landscape of twins with MS is shifted towards a monocyte population with increased expression of CCR2 and the GM-CSF receptor. **a**, Network visualization of unsupervised FlowSOM clusters in the myeloid compartment (beige); highlighted are the differential state nodes for twins with MS for which features appeared significantly different (red) and nodes of the manually annotated reference framework (green). The violin plots show the median expression levels of CCR2 and CD116 in monocytes within the differential state node from untreated twin pairs and all combined twins enrolled in the study; the bold horizontal line depicts the respective group mean and the dashed line indicates twinship. **b**, Mapping of diffcyt-generated myeloid cell clusters (right) on InfinityFlow data from PBMCs of a healthy

donor outside the cohort of twins with MS (left). The dot size in **a, b** corresponds to the population frequency among total leukocytes. cDC1s, conventional type 1 dendritic cells. **c**, Dot plot of CITE-seq data (a total of 21,043 cells) showing the expression profile of the top 15 differentially expressed genes for each cluster. The dot size corresponds to the fraction of cells within each cluster expressing the indicated transcript, and the colour indicates average expression. If not indicated, the differences between the experimental groups were statistically not significant ($P > 5\%$) using the moderated limma-trend method implemented in diffcyt, performing a false discovery correction according to the Benjamini–Hochberg approach.

healthy control sample outside the cohort of twins with MS²⁴. InfinityFlow combines traditional flow cytometry-based detection of hundreds of marker molecules, with machine learning analysis, facilitating the mapping of cellular subsets with single-cell resolution²⁵. We applied FlowSOM clustering to identify the canonical myeloid cell subsets of the peripheral blood and retrieved their surface proteome (Extended Data Fig. 2d). We next projected the differential state phagocyte node from the twin pairs with MS onto this map and identified their surface profile to be most similar to the classical monocyte and intermediate monocyte nodes (Fig. 2b, Extended Data Fig. 2e).

To gain further insight into the functional properties of the differential state phagocyte node, we performed CITE-seq of the myeloid compartment of monozygotic twin pairs discordant for MS and obtained similar populations as present in the CyTOF data (Extended Data Fig. 3a, Extended Data Table 1). CITE-seq confirmed that the phagocyte node characterized by the CD14^{dim}CD16⁺ phenotype showed transcriptional similarity to both classical monocytes and dendritic cells, as well as being characterized by the expression of *CIITA*, *ZEB2*, *JAK2* and *ITGAX* (Fig. 2c, Extended Data Fig. 3a, b). In addition, we observed a trend towards increased signalling activity of *CSF2RA* with increased expression of *CSF2RB*, *PRKACA* and *STAT5A* transcripts in twins with MS compared with their unaffected twin siblings (Extended Data Fig. 3c).

Exploring the monocyte reference nodes in proximity to the differential state node in the cohort of twins with MS, we further revealed that the frequency of non-classical monocytes among total myeloid cells was significantly reduced in twins with MS compared with their unaffected twin siblings. Again, this difference was more pronounced in untreated twin pairs (Extended Data Fig. 3d), and indeed treatment with dimethyl fumarate, fingolimod or glatiramer acetate significantly reduced the disparity between twins with MS and their unaffected twin siblings (Extended Data Fig. 3e). Finally, investigation of the transcriptional circuits, which differentiate twins with MS from unaffected twin siblings in the reference nodes of the myeloid compartment, revealed increased expression of *FKBP5*, *CCND3*, *PER1* and *IRAK3* in classical monocytes of twins with MS compared with their unaffected twin siblings (Extended

Data Fig. 3f, Supplementary Table 4). Moreover, we observed an overall reduction in the type 1 interferon gene signature in classical monocytes, dendritic cells and non-classical monocytes of twins with MS compared with twins without MS, providing a potential link towards the clinical efficacy of recombinant interferon- β therapy in MS²⁶ (Extended Data Fig. 3f, g, Supplementary Table 4).

In conclusion, twins with MS manifested a population shift in the myeloid compartment, away from non-classical monocytes and towards their inflammatory classical monocyte counterparts with a concomitant decrease in the type 1 interferon gene signature. A subpopulation of monocytes exhibited elevated expression of CCR2 and the GM-CSF receptor, indicative of a sensitization towards inflammatory stimuli.

T_H cells display increased CD25 in MS

The unbiased feature extraction revealed the most prominent perturbations between twins with MS and unaffected twins within the lymphocyte compartment (Fig. 1c, Extended Data Table 2). Within the innate lymphocyte compartment, we observed increased expression of CCR6 and reduced amounts of CXCR3 and CD25 in NK cells and lower levels of CCR6 expression in an NKT cell node of twins with MS (Extended Data Fig. 4a). Significantly lower expression of CXCR3 and higher expression of CD69 were found in a node of CD4⁺ effector memory T cells re-expressing CD45RA (T_{EMRA}) in twins with MS (Fig. 3a, Extended Data Fig. 4b).

However, the most dominant perturbations across our twin cohort were present in T_H cell nodes: we observed significantly higher expression of CD25, the IL-2 receptor high-affinity chain, in eight clusters phenotypically characterized as naive T_H cells that was accompanied by increased expression of CCR4 and decreased expression of CD127 in one node, respectively (Fig. 3a, Extended Data Fig. 4c). Although the overall expression of CD25 across naive T_H cells was relatively low, reducing the signal-to-noise ratio of the twin setup combined with barcoding and simultaneous acquisition revealed a robust increase in the expression of CD25 in twins with MS. This perturbation was the statistically most substantial change in the data-driven immune profiling of the cohort of twins with MS. CD25

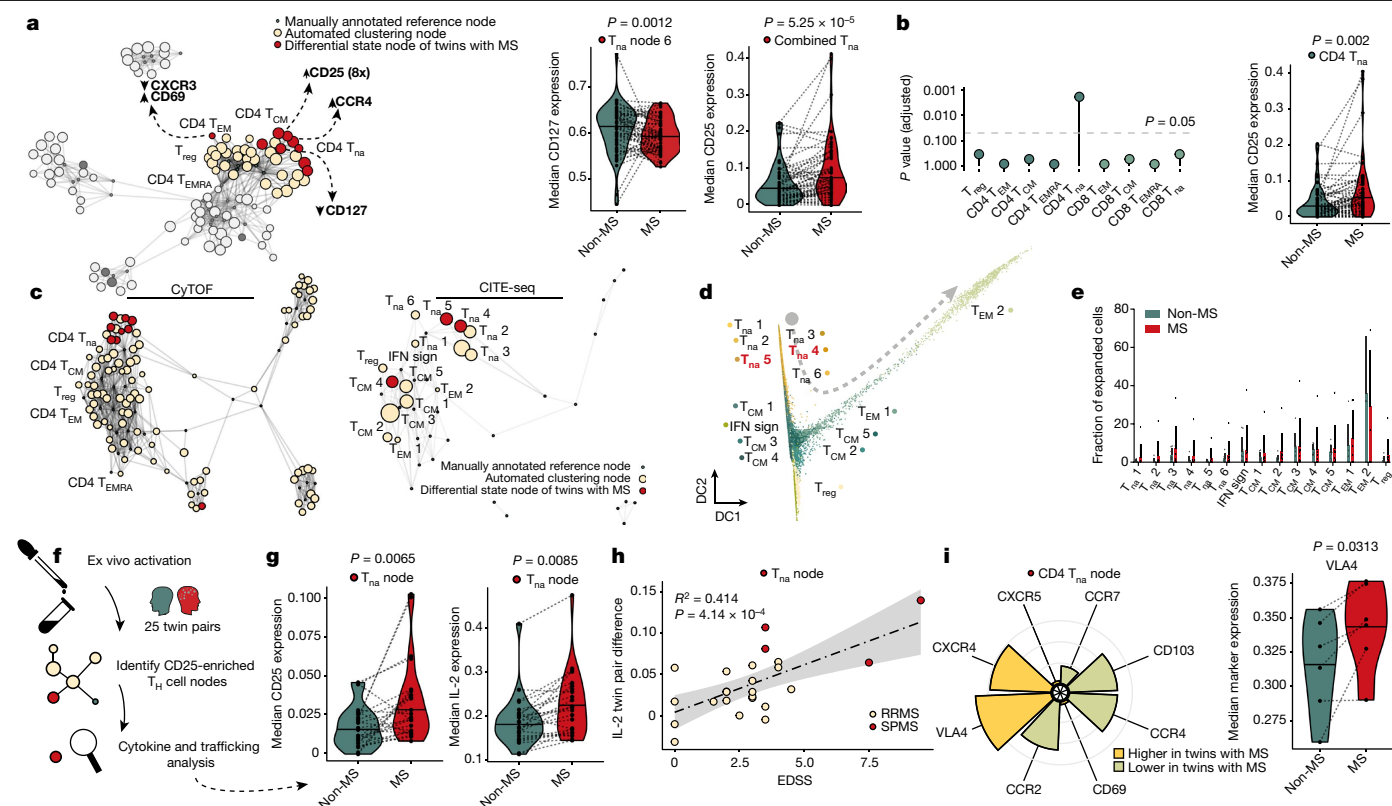


Fig. 3 | Transitional T_H cells of twins with MS display IL-2 hypersensitivity and elevated expression of brain-homing markers compared with unaffected twin siblings. **a**, Network visualizations of T_H cell clusters yielded by diffcyt analysis; highlighted are the differential state nodes for twins with MS for which features appeared significantly different (red, differential markers annotated in bold, 8x indicates number of differential nodes for the same feature) and nodes of the manually annotated reference framework (green). The violin plots show the median expression level for CD127 in cells within the differential state T_{na} 6 node and the median expression level for CD25 within the combined nodes for all twin pairs. **b**, Lollipop and violin plots showing the median expression levels of CD25 and the resulting P values across all T cell nodes in the reference framework of the naive T_H cell compartment for twin pairs discordant for MS ($n = 57$). **c**, Mapping of CITE-seq nodes (78,531 cells) from eight twin pairs discordant for MS onto the mass cytometry dataset. The red nodes indicate significantly increased surface expression of CD25 between twin pairs. **d**, Diffusion map showing the trajectory of T_H cells. The arrow indicates the direction of trajectory and nodes highlighted in red demonstrated increased CD25 expression in twins with MS. DC, diffusion coefficients. **e**, Bar graph showing the fraction of expanded T cell clones ($n > 2$ per clonotype) for twins discordant for MS (78,531 cells). **f**, PBMCs were activated in an antigen-independent manner, CD25-enriched T_H cell

represents a major genetic MS risk allele²⁷. It is therefore of great interest that the unbiased analysis of twin siblings discordant for MS, with a shared genetic makeup, identified CD25 expression as one of the main features of the differential state node in T_H cells in siblings with MS.

When we mapped the eight differential state naive T_H cell nodes characterized by elevated expression of CD25 onto the reference T_H cell nodes for twins with MS, we found that they were phenotypically similar (Extended Data Fig. 4c, d). Binning of the eight clusters into a single naive T_H cell node also further increased the pairwise significance of the inter-twin difference in CD25 expression (Extended Data Fig. 4e). Investigation of our manually annotated canonical T cell reference nodes revealed that the expression of CD25 was low in naive T_H cells. However, the expression of CD25 in naive T_H cells was still higher than cytotoxic T cells (Extended Data Fig. 4f). Although significant differences in the expression of CD25 between pairs of twins with MS and their unaffected twins were restricted to T_H cells displaying

nodes were identified and analysed with regards to their cytokine and the trafficking profile in twin pairs discordant for MS ($n = 25$). **g**, Violin plot showing the median expression levels of CD25 and IL-2 in cells within the differential state nodes in twins with MS for all twin pairs. **h**, Correlation between the inter-pair difference in median expression of IL-2 and expanded disability status scale (EDSS) of twins with MS. RRMS is shown in beige ($n = 21$), and secondary progressive MS (SPMS) is shown in red ($n = 4$). The dashed line indicates the smoothed conditional mean of the linear regression model with a 95% confidence interval in the shaded area. **i**, Effect size for the expression of the indicated trafficking molecules (left) and a violin plot showing the median expression level of VLA4 (right) in cells within the differential state T_H cell node of twin pairs discordant for MS ($n = 6$). The dot size of the networks corresponds to the population frequency among total leukocytes. In the violin plots, a bold horizontal line depicts the respective group mean and the dashed line indicates twinning. If not indicated, the differences between experimental groups were statistically not significant ($P > 5\%$) using the moderated limma-trend method implemented in diffcyt (**a**, left) or a two-sided paired non-parametric Wilcoxon signed-rank test (**a**, right, **b**, **g**, **i**), both applying a false discovery correction according to the Benjamini–Hochberg approach or a paired two-way analysis of variance (ANOVA) (**e**).

a naive phenotype, the most marked differences were limited to a very discrete set of T_H cell nodes (Fig. 3b, Extended Data Fig. 4g).

Correlation analysis of significantly dysregulated immune traits in the cohort of twins with MS revealed that CD25 features in naive T_H cells not only strongly correlated with each other but also with the expression of CD116 and CCR2 in monocytes, suggesting interconnectivity between these features during immunopathology (Extended Data Fig. 4h). Together, elevated expression of CD25 within a subgroup of phenotypically naive T_H cells represented the most pervasive immune alteration in circulating immune cells of twins with MS.

CD25 affects transitional T_H cells in MS

To further expose the cellular identity of the CD25 differential state T_H cell nodes in twins with MS, we performed CITE-seq, thereby retrieving

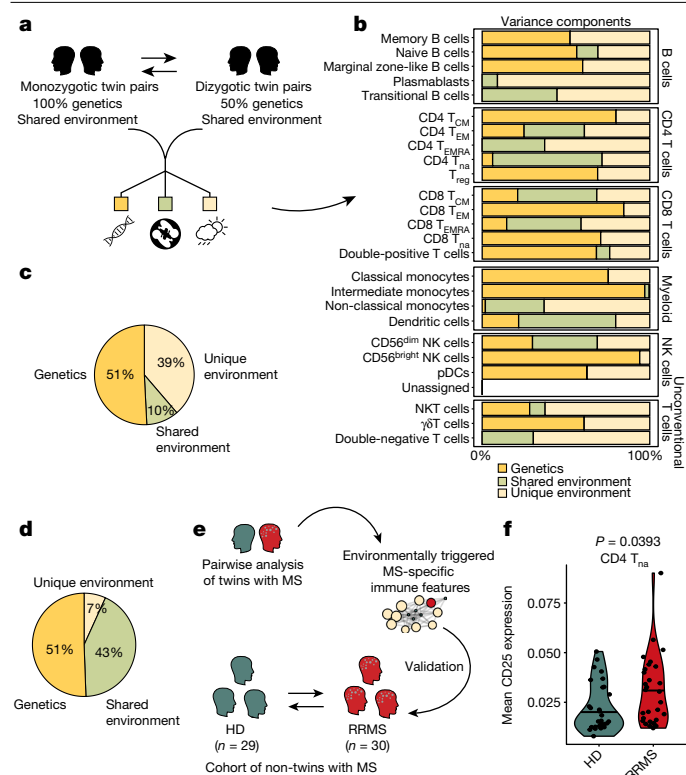


Fig. 4 | The expression of CD25 in naive T_H cells is regulated by genetic and early environmental factors and is increased in patients with MS in a cross-sectional validation cohort. **a**, Flow cytometry data of PBMCs from healthy monozygotic ($n = 21$) and dizygotic ($n = 22$) twin pairs age-matched to the cohort for twins with MS were used in a structural equation model to estimate the contribution of genetics and shared and non-shared environmental drivers on immune composition. **b–d**, Bar graphs (**b**) and pie charts (**c**, **d**) displaying the variance components for the populations in the manually annotated reference framework (**b**), the mean variance components across all detected immune subsets (**c**) and the variance components for the expression of CD25 in T_H cells displaying a naive phenotype (**d**). **e**, Data-driven analysis of twin pairs discordant for MS revealed MS-specific immune features that were validated within a cross-sectional validation cohort consisting of healthy donors (HD; $n = 29$) and patients with RRMS ($n = 30$). **f**, Violin plot showing the mean expression of CD25 in T_H cells displaying a naive phenotype of healthy donors and patients with RRMS. In the violin plots, the bold horizontal line depicts the respective group mean. If not indicated, differences between experimental groups were statistically not significant ($P > 5\%$) using a two-sided unpaired non-parametric Mann–Whitney–Wilcoxon test with a false discovery correction according to the Benjamini–Hochberg approach.

the single-cell transcriptome in combination with single-cell TCR and surface epitope information.

We identified six T_H cell clusters displaying a naive phenotype based on RNA (*CCR7*, *CD7*, *TCF7*, *LEF1* and *SELL*) and epitope expression (CD45RA⁺ and CD45RO^{low}) (Extended Data Fig. 5a). CITE-seq confirmed elevated surface expression of CD25 in two clusters of naive T_H cells (T_{na}4 and T_{na}5), and a cluster of central memory T (T_{CM}) cells (T_{CM}4) in twins with MS compared with unaffected twins (Extended Data Fig. 5b). These clusters mapped in close proximity to the differential CD25 nodes of the mass cytometry dataset of twins with MS, thereby validating the high-throughput discovery CyTOF approach using an independent single-cell technology (Fig. 3c, Extended Data Fig. 5c).

Although the differential state T_H cell nodes in twins with MS displayed a naive phenotype, the increased surface expression of CD25 in twins with MS was indicative of cellular activation. Indeed, the transcriptomes of the T_{na}4 and T_{na}5 clusters placed these cells distant from bona fide naive T_H cells based on pseudotime analysis and both clusters

started to downregulate naive markers (CD45RA and CCR7) while upregulating activation or memory-associated proteins (CD45RO and CD25) (Fig. 3d, Extended Data Fig. 5d). Next, we investigated whether the increased expression of CD25 in the transitional T_H cell nodes of twins with MS is accompanied by an expansion of TCR clones. The fact that there was no indication for clonal expansion across T_H cells of twins with MS compared with unaffected twin siblings does not support the notion that the observed activation is driven by a shared autoantigen (Fig. 3e, Extended Data Fig. 5e).

To uncover the underlying transcriptional circuits that drive the difference in CD25 expression, gene expression of transitional T_H cell nodes was analysed. Apart from transcripts involved in IL-2 signalling such as *CISH*, increased expression of transcripts associated with protein synthesis and proliferation (*EEF1A1*, *EEF1B2*, *EIF3L*, *PIK3IP1* and *TPT1*) was found in twins with MS compared with unaffected twins (Extended Data Fig. 5f, g, Supplementary Table 5). The transcript showing the most significant induction in twins with MS compared with unaffected twins was *TXNIP*, which encodes thioredoxin-interacting protein (also known as vitamin D₃ upregulated protein 1), a key mediator of the cellular antioxidant system that has been reported to regulate the responsiveness of T cells to IL-2 (ref. 28) (Extended Data Fig. 5g, Supplementary Table 5).

In summary, the unbiased single-cell transcriptome and epitope analysis of T_H cells across a selected group of twin pairs validated the specific increase of CD25 expression in a population of T_H cells in twins with MS initially captured by high-throughput mass cytometry. The transcriptome and pseudotime profiling further suggests that this population is within a transitional differentiation state with increased expression of genes related to protein biosynthesis and IL-2-induced proliferation.

Cytokine dysregulation in MS progression

As CD25 has a crucial function regulating T cell proliferation and activation²⁹, we next investigated the possible functional implications of increased expression of CD25 in the differential state transitional T_H nodes of twins with MS. Polyclonal stimulation of PBMCs from twin pairs discordant for MS using phorbol 12-myristate 13-acetate and ionomycin was followed by analysis of cytokines and trafficking-related markers (Fig. 3f). To anchor the CD25 differential state nodes in this additional dataset, a similar strategy as described above was applied. This revealed a T_H cell node displaying a naive phenotype characterized by increased expression of CD25 as well as its ligand IL-2 (Fig. 3g, Extended Data Fig. 5h, i). We observed a strong positive correlation between the size of inter-twin difference in IL-2 expression in the transitional T_H cell node and the severity of MS in the affected twin, assessed by the expanded disability status scale (Fig. 3h). Accordingly, the inter-twin difference in IL-2 expression was significantly higher in twin pairs in which the twin with MS had entered the more-advanced phase of disease progression known as secondary progressive MS, than twin pairs in which the affected twin was in a relapsing–remitting stage (RRMS) (Fig. 3h, Extended Data Fig. 5j).

To reveal early cytokine polarization across these transitional T_H cells, we gated on T_H cells that responded with IL-2 production to ex vivo reactivation (Extended Data Fig. 5k) and uncovered significantly lower production of IL-9 and higher production of IL-17A and IL-3 in twins with MS than in their unaffected twin sibling (Extended Data Fig. 5l, m). Both, IL-17A and IL-3 were reported to be secreted by encephalitogenic T cells in MS^{30,31}, suggesting that these transitional, CD25-expressing, IL-2-producing peripheral T_H cells are the precursors to fully encephalitogenic T_H cells found in CNS lesions in patients with MS. To investigate putative transcriptional networks that drive this early polarization fate, DoRothEA, a computational tool to assess transcription factor activities based on reported transcription factor–target interactions, was used. Significantly higher *STAT4* activity in both transitional T_H cell nodes that demonstrated increased expression of CD25 (T_{na}4 and T_{na}5) coincided with elevated *RELA* activity in T_{na}4 in twins with MS (Extended Data Fig. 5n, Supplementary Table 6). *RELA*, which encodes the NF-κB p65

subunit, has been shown to promote differentiation of IL-17-producing T cells³² and has been linked to MS in genome-wide association studies³³. Similarly, mice deficient in STAT4 were resistant to experimental autoimmune encephalomyelitis, an animal model of neuroinflammation³⁴.

MS T_H cells display a trafficking signature

To follow the hypothesis that the proliferative T_H cell node in twins with MS could give rise to encephalitogenic T cells, we next measured the expression of trafficking molecules in the differential state T_H cell node. Compared with cells from the unaffected twin siblings, the cells from twins with MS expressed significantly higher amounts of VLA4, which was further validated using our CITE-seq dataset (Fig. 3i, Extended Data Fig. 5o). VLA4 is required for leukocyte migration into the CNS³⁵ and is directly targeted by natalizumab, an approved therapy for MS³⁶. Finally, *CXCR4* transcripts were enriched in the two T_H cell nodes that demonstrated increased expression of CD25 (T_{na} 5 and T_{cm} 4) in twins with MS compared with unaffected twins, further supporting the CNS trafficking potential of the identified transitional T_H cells (Extended Data Fig. 5g).

Together, in the peripheral blood of twins with MS, a population of transitional T_H cells is expanding due to increased expression of CD25. The level of expansion, determined by the production of IL-2 after *ex vivo* activation, correlates with disease severity. Expanding cells in twins with MS displayed early signs of encephalitogenic polarization, possibly driven by STAT4 and NF- κ B activation, and demonstrated increased potential to traffic to the CNS.

CD25 in naive T_H cells is a heritable trait

Using the cohort of twins with MS, we were able to control for the genetic contribution to MS and thereby map immune alterations purely elicited by environmental cues. To extend this approach and estimate the respective effect of heritable, early environmental and late environmental influences on each immune trait in the reference framework, we next integrated flow cytometry data on PBMCs from a cohort of healthy monozygotic and dizygotic twin pairs (termed healthy twin cohort)¹⁷ that was age-matched to the cohort of twins with MS (Fig. 4a, Extended Data Table 1). Manual gating was used to match the cell populations from the healthy twin cohort to those in the reference framework of the cohort of twins with MS (Extended Data Fig. 6a). Using OpenMx, a structural equation model commonly utilized in twin studies, the extent to which each immune trait was modulated by genetic, early-shared or later unique environmental sources of variance during adulthood was estimated. The predominant sources of variance in immune cell frequencies differed between individual leukocyte subsets; although variance in frequency of CD56^{bright} NK cells and classical monocytes had a strong genetic component, plasmablasts and CD4⁺ T_{EMRA} cells were primarily influenced by unique environmental factors (Fig. 4b). Similar to previous reports^{17,37,38}, we observed that approximately 50% of the variance across all identified immune populations was attributable to genetic influences, 40% to unique environmental factors and 10% to a shared early environment (Fig. 4c). Next, this method was applied to identify the influences associated with modulation of CD25 expression in T_H cells overlapping with a naive phenotype. In naive T_H cells, 51% of the variance in CD25 expression was regulated by genetics, 43% by shared (and thus early childhood) environmental factors and only 7% by unique environmental drivers (Fig. 4d, Extended Data Fig. 6b). Accordingly, analysis of the cohort of twins with MS eliminated 93% of the variance caused by heritable and early-childhood factors, thereby isolating disease-associated immune alterations that are unaffected by confounding variables.

CD25 signature in non-twin patients with MS

Although the expression of CD25 in phenotypically naive T_H cells is largely regulated through heritable influences, the elevated expression

of CD25 in transitional, quasi-naive T_H cells in MS is a disease-specific effect driven by unique environmental influences. To further solidify this concept, we analysed the expression of CD25 in a cross-sectional validation cohort of 30 untreated patients with RRMS and 29 genetically unrelated healthy donors¹⁴ (cohort of non-twins with MS) (Fig. 4e, Extended Data Table 1). Here the frequencies of canonical immune cell subsets were again comparable between patients with MS and non-familially related healthy donors¹⁴ (Extended Data Fig. 6c, d). Despite multiple sources of variation and the absence of enrichment for a familial MS susceptibility in controls of the cross-sectional cohort of non-twins with MS, we could observe that unfractionated naive T_H cells of patients with RRMS expressed higher levels of CD25 than healthy donors (Fig. 4f).

In conclusion, we show that elevated expression of CD25 in T_H cells that display a naive phenotype is part of the heritable (cross-sectional) and non-heritable (twins) immune signature of MS.

Discussion

Here we used deep immune profiling of monozygotic twin pairs discordant for MS using two highly complementary single-cell technologies to begin to explain how two people sharing the same genetic predisposition for this disease can demonstrate such divergent clinical phenotypes. Data-driven analysis revealed that mainly the T_H cell and, to a lesser extent, the myeloid cell compartment were phenotypically altered in the systemic immune compartment of twins with MS. This confirmed and refined the findings of previous MS immunophenotyping studies, which could not control for the effects of genetic predisposition^{13,14}. The observation that some of the findings revealed in this study have previously been reported in cross-sectional studies of MS (for example, increase in CCR2 expression in monocytes³⁹), whereas other features demonstrated opposite trends of what has previously been described⁴⁰, highlights the importance of discerning genetic predisposition from environmentally induced alterations in MS.

Compared with their unaffected twin siblings, the twins with MS exhibited a shift in their circulating monocyte compartment away from tissue-patrolling non-classical monocytes, and towards inflammatory monocytes. Analogously, ref. ⁴¹ reported that non-classical monocytes enter and patrol the CNS during the steady state, but are reduced and superseded by classical monocytes in the peripheral blood and cerebrospinal fluid of patients with RRMS. The role of inflammatory monocytes has been investigated intensively in the context of experimental autoimmune encephalomyelitis. Genetic ablation models have demonstrated that CCR2-expressing monocytes are the main executors of CNS immunopathology in this model, in which they sense T-cell-derived GM-CSF and adopt a pathological CNS-specific transcriptional signature that leads to tissue damage and neurological deficits^{21,23,42}. We found that expression of both CCR2 and the GM-CSF receptor (CD116) was consistently increased in a subset of circulating monocytes in twins with MS. It is thus conceivable that T cell-orchestrated cytokine production leads to the observed phenotypic changes in the monocyte compartment, which may appear as an effect of the ongoing inflammation and ultimately contribute to tissue destruction in the CNS.

The most consistent pattern emerging from the pairwise analysis of monozygotic twins discordant for MS was significantly higher expression of CD25 in a population of T_H cells in patients with MS and their hyper-proliferative state. This exclusively affected a cluster of T_H cells transitioning from a naive to a memory or effector phenotype. Altered expression of CD25 in transitional T_H cells, predominantly under the influence of shared genetic and early environmental factors, highlights the crucial role of genetic predisposition in initiating MS. Despite eliminating the majority of heritable variance using the twin setting, the increase in the expression of CD25 in twins with MS still appeared as the most consistent immune dysregulation in the cohort of twins with MS. It is thus conceivable that polygenic risk variants, including single-nucleotide polymorphisms in the *IL2RA* gene^{27,43}, confer genetic susceptibility that

facilitates disease initiation by environmental challenges (for example, Epstein–Barr virus infection). Analogously, gene–phenotype correlations using clinically healthy high-risk allele carriers have revealed that the MS risk single-nucleotide polymorphism rs2104286 (in the *IL2RA* gene) resulted in increased expression of CD25 exclusively in naive T_H cells⁴⁴. The postulated hypersensitivity for IL-2 in T_H cells in patients with MS has been shown to induce secretion of GM-CSF in high-risk allele carriers¹¹, which represents a hallmark of CNS inflammation in both experimental autoimmune encephalomyelitis and human MS^{14,45}.

Functionally, the increased expression of CD25 and IL-2 responsiveness of transitional T_H cells from patients with MS may provide the substrate for pathogenic CD4⁺ effector memory T cells in the inflamed CNS, as previously proposed^{11,46}. This hypothesis aligns with the cytokine profile and expression of CNS-homing molecules such as CXCR4 and VLA4 in these T_H cells. Alternatively, homeostatically expanding CD25^{high} naive T_H cells⁴⁷ found in the peripheral blood of individuals with MS might contribute, as bystanders, to an antigen-specific immune response occurring in secondary lymphoid organs or the CNS, amplifying cytokine production up to the threshold required to start or maintain the neuroinflammatory cascade of MS. In support of this notion, IL-2-mediated STAT5 signalling is sufficient to induce the production of GM-CSF in naive human T_H cells^{11,46}. Accordingly, in the identified transitional T_H cell node, the increased expression of CD25 was accompanied by increased IL-2 in twins with MS and correlated with disease severity. Together with the changes in the composition of the circulating mononuclear phagocyte population, this indicates that a cytokine dysregulation characterized by activation of the IL-2RA–IL-2–GM-CSF axis, triggered by a unique environmental challenge, may well be the immunological substrate for disease initiation and/or progression in MS.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-04419-4>.

- Wallin, M. T. et al. Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 269–285 (2019).
- O’Gorman, C., Lin, R., Stankovich, J. & Broadley, S. A. Modelling genetic susceptibility to multiple sclerosis with family data. *Neuroepidemiology* **40**, 1–12 (2012).
- Sawcer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
- Simpson, S., Blizard, L., Otahal, P., Van Der Mei, I. & Taylor, B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **82**, 1132–1141 (2011).
- Nielsen, N. M. et al. Neonatal vitamin D status and risk of multiple sclerosis: a population-based case–control study. *Neurology* **88**, 44–51 (2017).
- Gardener, H. et al. Prenatal and perinatal factors and risk of multiple sclerosis. *Epidemiology* **20**, 611–618 (2009).
- Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372**, 1502–1517 (2008).
- Westerlind, H. et al. Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. *Brain* **137**, 770–778 (2014).
- Patsopoulos, N. A. et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).
- International Multiple Sclerosis Genetics Consortium. Low-frequency and rare-coding variation contributes to multiple sclerosis risk. *Cell* **175**, 1679–1687.e7 (2018).
- Hartmann, F. J. et al. Multiple sclerosis-associated IL2RA polymorphism controls GM-CSF production in human T_H cells. *Nat. Commun.* **5**, 5056 (2014).
- Smets, I. et al. Multiple sclerosis risk variants alter expression of co-stimulatory genes in B cells. *Brain* **141**, 786–796 (2018).
- Jelcic, I. et al. Memory B cells activate brain-homing, autoreactive CD4⁺ T cells in multiple sclerosis. *Cell* **175**, 85–100.e23 (2018).
- Galli, E. et al. GM-CSF and CXCR4 define a T helper cell signature in multiple sclerosis. *Nat. Med.* **25**, 1290–1300 (2019).
- Tzartos, J. S. et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. *Am. J. Pathol.* **172**, 146–155 (2008).
- Duddy, M. et al. Distinct effector cytokine profiles of memory and naive human B cell subsets and implication in multiple sclerosis. *J. Immunol.* **178**, 6092–6099 (2007).

- Roederer, M. et al. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell* **161**, 387–403 (2015).
- Brodin, P. et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47 (2015).
- Vogel, D. Y. S. et al. Macrophages in inflammatory multiple sclerosis lesions have an intermediate activation status. *J. Neuroinflammation* **10**, 35 (2013).
- Frischer, J. M. et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain* **132**, 1175–1189 (2009).
- Mildner, A. et al. CCR2^{hi} Ly-6C^{hi} monocytes are crucial for the effector phase of autoimmunity in the central nervous system. *Brain* **132**, 2487–2500 (2009).
- King, I. L., Dickendesher, T. L. & Segal, B. M. Circulating Ly-6C⁺ myeloid precursors migrate to the CNS and play a pathogenic role during autoimmune demyelinating disease. *Blood* **113**, 3190–3197 (2009).
- Croxford, A. L. et al. The cytokine GM-CSF drives the inflammatory signature of CCR2⁺ monocytes and licenses autoimmunity. *Immunity* **43**, 502–514 (2015).
- Dutertre, C. A. et al. Single-cell analysis of human mononuclear phagocytes reveals subset-defining markers and identifies circulating inflammatory dendritic cells. *Immunity* **51**, 573–589.e8 (2019).
- Becht, E. et al. InfinityFlow: high-throughput single-cell quantification of 100s of proteins using conventional flow cytometry and machine learning. Preprint at <https://doi.org/10.1101/2020.06.17.152926> (2020).
- Marrie, R. A. & Rudick, R. A. Drug insight: interferon treatment in multiple sclerosis. *Nat. Clin. Pract. Neurol.* **2**, 34–44 (2006).
- The International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–862 (2007).
- Ahsan, M. K. et al. Loss of interleukin-2-dependency in HTLV-1-infected T cells on gene silencing of thioredoxin-binding protein-2. *Oncogene* **25**, 2181–2191 (2006).
- Boyman, O. & Sprent, J. The role of interleukin-2 during homeostasis and activation of the immune system. *Nat. Rev. Immunol.* **12**, 180–190 (2012).
- Lee, P. W., Xin, M. K., Pei, W., Yang, Y. & Lovett-Racke, A. E. IL-3 is a marker of encephalitogenic T cells, but not essential for CNS autoimmunity. *Front. Immunol.* **9**, 1255 (2018).
- Cao, Y. et al. Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. *Sci. Transl. Med.* **7**, 287ra74 (2015).
- Ruan, Q. et al. The Th17 immune response is controlled by the Rel–RORγ–RORγT transcriptional axis. *J. Exp. Med.* **208**, 2321–2333 (2011).
- Hussman, J. P. et al. GWAS analysis implicates NF-κB-mediated induction of inflammatory T cells in multiple sclerosis. *Genes Immun.* **208**, 2321–2333 (2016).
- Chitnis, T. et al. Effect of targeted disruption of STAT4 and STAT6 on the induction of experimental autoimmune encephalomyelitis. *J. Clin. Invest.* **108**, 739–747 (2001).
- Schneider-Hohendorf, T. et al. VLA-4 blockade promotes differential routes into human CNS involving PSGL-1 rolling of T cells and MCAM-adhesion of TH17 cells. *J. Exp. Med.* **211**, 1833–1846 (2014).
- Rudick, R. A. et al. Natalizumab plus Interferon beta-1a for relapsing multiple sclerosis. *N. Engl. J. Med.* **354**, 911–923 (2006).
- de Craen, A. J. M. et al. Heritability estimates of innate immunity: an extended twin study. *Genes Immun.* **6**, 167–170 (2005).
- Gerdes, L. A. et al. Immune signatures of prodromal multiple sclerosis in monozygotic twins. *Proc. Natl Acad. Sci. USA* **117**, 21546–21556 (2020).
- Chuluundorj, D., Harding, S. A., Abernethy, D. & La Flamme, A. C. Expansion and preferential activation of the CD14⁺ CD16⁺ monocyte subset during multiple sclerosis. *Immunol. Cell Biol.* **92**, 509–517 (2014).
- Gjelstrup, M. C. et al. Subsets of activated monocytes and markers of inflammation in incipient and progressed multiple sclerosis. *Immunol. Cell Biol.* **96**, 160–174 (2018).
- Waschbisch, A. et al. Pivotal role for CD16⁺ monocytes in immune surveillance of the central nervous system. *J. Immunol.* **196**, 1558–1567 (2016).
- Spath, S. et al. Dysregulation of the cytokine GM-CSF induces spontaneous phagocyte invasion and immunopathology in the central nervous system. *Immunity* **46**, 245–260 (2017).
- Weber, F. et al. IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. *Genes Immun.* **9**, 259–263 (2008).
- Dendrou, C. A. et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nat. Genet.* **41**, 1011–1015 (2009).
- Codarri, L. et al. RORγ3T drives production of the cytokine GM-CSF in helper T cells, which is essential for the effector phase of autoimmune neuroinflammation. *Nat. Immunol.* **12**, 560–567 (2011).
- Noster, R. et al. IL-17 and GM-CSF expression are antagonistically regulated by human T helper cells. *Sci. Transl. Med.* **6**, 241ra80 (2014).
- Pekalski, M. L. et al. Postthymic expansion in human CD4 naive T cells defined by expression of functional high-affinity IL-2 receptors. *J. Immunol.* **190**, 2554–2566 (2013).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Sample selection for the cohort of twins with MS

The cohort of twins with MS is part of the MS TWIN STUDY and represents a cohort of monozygotic twins with discordance for MS and is located at the Institute of Clinical Neuroimmunology at the LMU Klinikum Munich, Germany. Recruitment started in May 2012 and is still ongoing; samples used in the present study were collected up to May 2020.

Inclusion criteria for study participation were met if in one twin of a monozygotic twin pair a diagnosis of MS according to the revised McDonald criteria^{48,49} was established, whereas the twin sibling was clinically healthy. Exclusion criteria were infection as well as treatment with antibiotics or high-dose intravenous glucocorticosteroids within 3 months before sampling. Monozygotic twin pairs clinically discordant for MS ($n = 61$) visited the outpatient department at the Institute of Clinical Neuroimmunology at the LMU Klinikum Munich for a detailed interview, neurological examination, blood sampling and MRI investigations (in a proportion of twins only). To confirm a diagnosis of MS, medical records including MRI scans were obtained and reviewed (Extended Data Table 1, Supplementary Tables 1, 2).

As the current disease-modifying treatment at the time of blood sampling is known to have a strong effect on the peripheral immune signature³⁸, we selected a subgroup, in which the twin with MS had not received treatment at the time of blood sampling ($n = 20$) for further analyses (Extended Data Table 1, Supplementary Tables 1, 2). The expanded disability status scale (EDSS) was used as a measure of disease severity in twins with MS⁵⁰.

The MS TWIN STUDY was approved by the local ethics committees of the Ludwig-Maximilians-University of Munich (ethics approval project number 267-13). All participants gave written informed consent, according to the principles of the Declaration of Helsinki.

Blood sampling and PBMC preparation for the cohort of twins with MS

Blood samples of study participants of the MS TWIN STUDY were collected in EDTA-containing tubes. To exclude sample collection bias, blood samples were drawn from each twin pair before meals and at the same time on the same day. PBMCs were isolated as described before by density gradient centrifugation with Lymphoprep (STEMCELL technologies) and cryopreserved in liquid nitrogen using serum-free cryopreservation medium (CTL-Cryo ABC Media Kit, Immunospot) in concentrations of 1×10^7 cells per ml.

Ex vivo activation of PBMCs in the cohort of twins with MS

To measure cytokine expression by PBMC from the cohort of twins with MS, cells from 25 twin pairs (19 treated pairs and 6 untreated pairs) were activated in an antigen-independent manner as described previously⁵¹. In brief, leukocytes were taken from liquid nitrogen storage and thawed in a water bath at 37 °C. Cells were resuspended in cell culture medium (RPMI-1640, 10% FCS (Biochrom), $1 \times$ L-glutamine and $1 \times$ penicillin-streptomycin (both Life Technologies)) supplemented with $1:10,000$ benzocaine (Sigma-Aldrich), then centrifuged ($350g$ for 7 min at 24 °C) and washed twice with cell culture medium. Samples subsequently underwent antibody labelling for mass cytometry, or, in the case of intracellular cytokine detection, were incubated overnight at 37 °C and 5% CO₂, before stimulation with 50 ng ml^{-1} phorbol 12-myristate 13-acetate (Sigma-Aldrich) and 500 ng ml^{-1} ionomycin (Sigma-Aldrich) in the presence of $1 \times$ brefeldin A and $1 \times$ monensin (both BD Biosciences) for 4 h at 37 °C. Cells underwent surface marker antibody labelling, fixation, permeabilization and intracellular cytokine antibody labelling as described below.

Antibodies

Antibodies used in mass cytometry experiments were either purchased already heavy-metal-conjugated (Fluidigm) or were conjugated in-house using the MaxPar X8 chelating polymer kit (Fluidigm)

following the manufacturer's instructions. Antibody clones, corresponding heavy metal tags and suppliers are summarized in Supplementary Table 3. Antibodies used in CITE-seq experiments were purchased pre-conjugated and are summarized in Supplementary Table 7.

Live-cell barcoding for mass cytometry

To eliminate technical variability during sample processing and data acquisition, a restricted combinatorial 9-choose-3 live-cell barcoding strategy was applied, as described previously¹⁴. To achieve this, anti-CD45 monoclonal antibodies (mAbs; BioLegend) were conjugated using MaxPar X8 polymers (Fluidigm) and palladium (¹⁰⁴Pd, ¹⁰⁵Pd, ¹⁰⁶Pd, ¹⁰⁸Pd and ¹¹⁰Pd), indium (¹¹³In and ¹¹⁵In; all from Trace Sciences International) and tantalum (¹⁸¹Ta; Sigma) isotopes. In addition, Y89-conjugated anti-CD45 mAbs (Fluidigm) were used. Twin pairs were randomized and PBMCs of twins with MS and unaffected twin siblings were barcoded in two batches before acquisition during two independent mass cytometry runs. After sample thawing and/or ex vivo activation cells were labelled with heavy-metal-tagged CD45 antibodies at 37 °C for 25 min in cell labelling medium (CLM; RPMI-1640, 4% FCS) on an orbital shaker (500 rpm). Barcoded samples were washed twice in CLM and combined into a single-reaction vessel for surface marker and/or cytokine detection.

Surface and intracellular cytokine detection by mass cytometry

After barcoding, the sample convolute was labelled in 400 μ l CLM containing the antibody mix for surface marker detection (Supplementary Table 3) on an orbital shaker (500 rpm) at 37 °C for 40 min. Cisplatin (2.5 μ M in PBS; Fluidigm) was added for 2 min on ice to enable live/dead cell discrimination and the reaction was stopped by adding 2% FCS in PBS and incubating for 2 min on ice.

For the detection of transcription factors, the barcoded sample convolute was fixed and permeabilized for 40 min at 4 °C in 1X FXP3 fixation/permeabilization buffer (BioLegend) and was washed in permeabilization buffer (PBS, 0.5% saponin, 2% bovine serum albumin (BSA) and 0.01% sodium azide (all Sigma-Aldrich)). Labelling was performed in 400 μ l permeabilization buffer containing the antibody mix for 1 h at 4 °C.

In case of intracellular cytokine detection, the sample convolute was fixed in 1.6% paraformaldehyde (Electron Microscopy Sciences) for 1 h at 4 °C and washed with permeabilization buffer. The sample was labelled with antibodies recognizing intracellular cytokines in 400 μ l permeabilization buffer for 1 h at 4 °C.

For both surface and nuclear antigen detection and intracellular cytokine detection, the labelled sample convolute was washed and incubated in 1X iridium intercalator solution (Fluidigm) at 4 °C overnight. The sample convolute was washed twice with PBS and twice with MaxPar water (Fluidigm) and following data were acquired.

Mass cytometry data acquisition and preprocessing

Data were acquired on a CyTOF 2.1 mass cytometer (Fluidigm) with daily instrument quality control and tuning. Acquisitions from two independent CyTOF runs, each containing both of the twin siblings, were normalized using five-element beads (Fluidigm)⁵². To monitor potential batch effects, each independent run contained two normalization control samples that were present in both runs. Manual gating using FlowJo (TreeStar) was applied to identify live single cells in the sample convolute based on event length, centre, width, DNA (¹⁹¹Ir and ¹⁹³Ir) and live/dead (¹⁹⁵Pt) channels. Following this, the sample convolute was debarcoded by utilizing Boolean gates of cells exclusively bearing three barcodes to prevent barcode mis-identification and facilitate doublet exclusion. Mass cytometry data were transformed in the R environment using an inverse hyperbolic sine (arcsine) function with varying cofactors to account for labelling variability between individual markers. To remove residual batch effects between the two independent mass cytometry runs, individual markers of the sample

Article

convolute were aligned by modifying the arcsine cofactor to achieve the same mean in labelling intensity for both normalization controls. A marker-based percentile normalization was applied using the 99.9th percentile of the transformed dataset. Analogously, individual cytokine positivity was determined by calculating the 99th percentile of the residual cytokine intensity using an unstimulated control.

Algorithm-guided analysis of the cohort of twins with MS

The entire analysis was carried out in the statistical programming environment R in RStudio and Visual Studio Code (Microsoft). UMAPs were computed using the *umap* package with default parameters⁵³. For the generation of the reference framework, FlowSOM was applied⁵⁴ in a similar manner as described previously^{51,55}. In brief, 100 clusters of the combined dataset were generated and metaclustering was performed based on the elbow criterion determined by the ConsensusClusterPlus package. Resulting metaclusters were manually merged and annotated based on the median expression profile of individual metaclusters and localization on the UMAP. After initial clustering yielded the main populations such as CD4⁺ T cells, CD8⁺ T cells, B cells and myeloid cells, each main population was iteratively clustered into subpopulations that depicted the nodes of the reference framework.

The data-driven analysis was carried out using the *diffcyt* package⁵⁶. For this, markers were separated into cell-state and cell-type markers (Supplementary Table 3) and the differential state *diffcyt* analysis was carried out using a design matrix considering the paired design. Significant immune features between all twins with MS and unaffected twin siblings were extracted using the moderated *limma*-trend method implemented in *diffcyt*, applying a false discovery correction according to the Benjamini–Hochberg approach⁵⁷. To exclude treatment-induced immune alterations, a second filter was applied to identify significant immune features between only untreated twins with MS and unaffected twin siblings. Accordingly, immune features were extracted that fulfilled both conditions: (1) they appeared significantly different between all twins with MS and unaffected twin siblings, and (2) were not elicited by disease-modifying therapy of the twin with MS.

Network visualizations to map *diffcyt*-generated clusters onto the reference framework or to map different mass cytometry panels, InfinityFlow data or CITE-seq data were generated using modifications of the *grappolo* and *vite* packages of the Scaffold framework⁵⁸. Resulting force-directed graphs were rearranged using the Fruchterman–Reingold and Kamada–Kawai algorithms implemented in the *igraph* package and modified and visualized using *ggraph*. Heatmaps were drawn using the *heatmap* package. Correlograms were generated using the *Hmisc* and *corrplot* packages. All remaining plots were drawn using *ggplot2*.

Mapping of myeloid cell populations using InfinityFlow

To fine-map the differential state phagocyte node, an extended myeloid reference framework was generated using the InfinityFlow data of healthy PBMCs containing 14 backbone markers and 332 predicted markers as published by ref.²⁴. To generate the extended myeloid reference nodes, FlowSOM clustering was applied as described above using the backbone markers for clustering. The InfinityFlow dataset contained an overlap of 32 markers with the mass cytometry dataset of twins with MS that facilitated the mapping of the *diffcyt*-differential state node on the extended reference nodes using the *grappolo* and *vite* packages as described above.

Sorting of twin samples and single-cell sequencing

Frozen PBMC samples from eight twin pairs with MS (16 samples; four pairs analysed by both CyTOF and CITE-seq and four additional pairs exclusively by CITE-seq) plus two additional healthy samples (total of 18 samples) were thawed quickly and washed twice with 1% BSA in PBS at 4 °C, followed by centrifugation at 300g for 10 min. The PBMC pellet was stained with Human TruStain FcX Fc Blocking Reagent (BioLegend) for 10 min and further stained with Fixable Viability Dye APC-eFluor

780, anti-CD3-AF700 (clone OKT3, Invitrogen), anti-CD4-pacific blue (clone S3.5, Invitrogen), anti-CD11c-PE (clone BUI5, BioLegend), and TotalSeq-C antibodies (Supplementary Table 7) for 30 min on ice. CD4⁺ T cells were sort purified as a singlet, live, CD3⁺ and CD4⁺, on a FACS Aria Fusion (BD). Simultaneously, a second population was sorted as a singlet, live, CD3⁻ and CD11c⁺.

Sorted cells were washed in 0.04% BSA in PBS. Approximately 25,000 cells per population per sample were loaded on a 10x chip and run onto the 10x Chromium controller using Chromium NextGEM Single Cell V(D) J Reagent kits v1.1 with Feature Barcoding technology for Cell-Surface Protein (10x Genomics) according to the manufacturer's protocol. Gene expression, TCR enrichment and cell-surface protein expression were multiplexed using individual Chromium i7 Sample Indices. Gene expression and TCR enrichment libraries were sequenced on NovaSeq S4 Flowcells using 150-bp paired-end reads and 8 bp for the i7 index, aiming for 50,000 reads per cell for gene expression and 5,000 reads per cell for TCR enrichment. Cell-surface protein expression libraries were sequenced on NovaSeq S1 Flowcells using 50-bp paired-end reads and 8 bp for the i7 index, aiming for 10,000 reads per cell.

Single-cell sequencing data processing

Cell Ranger software (10x Genomics, v.6.1) was used to demultiplex samples, process raw data, align reads to the GRCh38 human reference genome and summarize unique molecular identifier (UMI) counts. Filtered gene-barcode and cell-surface protein expression-barcode matrices that contained only barcodes with UMI counts that passed the threshold for cell detection were used for further analysis. Then, we processed the filtered UMI count matrices using the R package Seurat (version 4.0.3)⁵⁹. Cells that expressed fewer than 500 genes and/or >15% mitochondrial reads, and genes expressed in fewer than three cells were removed from the count matrix. After quality control, only raw gene counts in high-quality singlets were submitted to: log-normalization; identification of high-variable genes by using the *vst* method; scaling; and regression against the number of UMIs and mitochondrial RNA content per cell. We applied an unbiased calculation of the *k*-nearest neighbours, generated the neighbourhood graph and embedding using UMAP. Differentially expressed genes between each cluster and all other cells were calculated using the FindAllMarkers function. Annotation of Seurat clusters was manually curated using a combination of upregulated genes for each cluster and visual inspection of key markers using UMAP visualization.

After initial cluster annotation, we subsetted all clusters containing myeloid cells and reanalysed this subset. After subsetting, integration using reciprocal principal component analysis was performed to remove batch effects and the integrated assay was used for principal component analysis and unsupervised clustering. Seurat subclusters were annotated using a combination of canonical protein and mRNA markers. Similar to myeloid cells, only CD4⁺ T cells in which a TCR clonotype was detected, were subsetted and reanalysed. Single-cell TCRs were computed from the TCR enrichment sequencing data using Cell Ranger vdj pipeline (10x Genomics, v.6.1). CD4⁺ T cells containing more than two β -chains were removed.

Only samples with data for both the sibling with MS and the unaffected sibling were used for downstream analysis: seven and eight twin pairs for myeloid and T_H cell populations, respectively.

Downstream analysis of CITE-seq data

Differentially expressed genes between MS-affected and unaffected twin siblings were computed using a logistic regression model using the twinship as a latent variable. Mapping of CITE-seq data onto the mass cytometry data was accomplished by arcsine transformation of the raw counts obtained for each surface marker followed by percentile normalization similarly as performed for mass cytometry data. Resulting transformed and normalized data were combined with mass cytometry data to create a cellular network using the *grappolo* and *vite* packages as

described above. Transcription factor regulon activity was inferred based on the gene expression levels of its targets using DoRothEA⁶⁰. Trajectory and pseudotime were computed based on the corresponding UMAP using Monocle 3 (ref. ⁶¹). Subsequently, the Seurat object was converted into the .h5ad format for further trajectory analysis and calculation of diffusion maps using the SCANPY analysis framework implemented in Python⁶². Surface marker expression along pseudotime was smoothed using a generalized additive model from the mgcv package.

Variance component analysis using the healthy twin cohort

To dissect genetic, early shared environmental and unique environmental sources of variance for frequencies of immune populations or median marker expressions within immune subsets, publicly available flow cytometry data from healthy monozygotic and dizygotic twin pairs were accessed¹⁷. Healthy monozygotic ($n = 21$) and dizygotic ($n = 22$) twin pairs were selected to be age-matched to the twin pairs in the cohort of twins with MS (Extended Data Table 1). A compensation matrix of the flow cytometry data was corrected using FlowJo (TreeStar) and populations matching the immune subsets of the reference framework in the cohort of twins with MS were retrieved using manual gating. Resulting immune subset frequencies or median marker expressions were imported into R. Variance components were estimated in a two-group Cholesky twin model using the umxACE function of the umx package applying default parameters—a framework commonly applied to twin studies^{63,64}.

Validation in the cross-sectional MS cohort

For the validation of the findings in the cohort of twins with MS, a publicly available cross-sectional cohort of non-twins with MS measured by mass cytometry was accessed¹⁴ (Extended Data Table 1). The analysis was carried out using exclusively untreated patients with RRMS. A main population and T_H cell subpopulation label for each cell was provided in the publicly available dataset.

Statistical analysis

Immune cell frequencies and median marker expressions were compared using the unpaired non-parametric Mann–Whitney–Wilcoxon test or the paired non-parametric Wilcoxon signed-rank test implemented in the stats package with a false discovery correction according to the Benjamini–Hochberg approach⁵⁷. For the validation cohort, the per-patient mean of CD25 expression was compared between patients with RRMS and health donors as a statistical parameter that is sensitive to outliers and could thus reflect variable CD25 expression by only a subset of naive T_H cells. Mass cytometry analysis was carried out in two independent runs and was not repeated due to limited precious sample material. CITE-seq analysis was carried out once. Gene expression, module scores and transcription factor activity between cells of twins with MS and unaffected twin siblings were compared using a logistic regression model with the twinship as a latent variable and comparing each model to a null model with a likelihood ratio test and applying a Bonferroni correction. Linear regression analysis was performed using the base lm function. Inter-twin pair effect sizes were computed using the Wilcoxon two-sample paired signed-rank test implemented in the rstatix package.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw mass cytometry data can be accessed at <https://doi.org/10.17632/fzs5ph5p8s.1>. CITE-seq data are available at <https://doi.org/10.17632/278fy5m2yj.2>. Publicly available flow cytometry data of healthy monozygotic and dizygotic twin pairs¹⁷ were accessed at <http://www.tinyurl.com/twinsFACSdata>. The publicly available cross-sectional

MS mass cytometry dataset for non-twins¹⁴ was accessed at <http://flowrepository.org/experiments/2166/>.

Code availability

The code for the mass cytometry analysis and the variance component analysis of the healthy monozygotic and dizygotic twin pairs is available at <https://github.com/florianingelfinger/MStwins>. The code for the CITE-seq analysis of the cohort for twins with MS can be accessed at https://github.com/beltranLab/twin_study_Nature_2021.

- Thompson, A. J. et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* **17**, 162–173 (2018).
- Polman, C. H. et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* **69**, 292–302 (2011).
- Bushnik, T. in *Encyclopedia of Clinical Neuropsychology* (eds Kreutzer, J. S., DeLuca, J. & Caplan, B.) https://doi.org/10.1007/978-3-319-57111-9_1805 (Springer, 2018).
- Hartmann, F. J. et al. High-dimensional single-cell analysis reveals the immune signature of narcolepsy. *J. Exp. Med.* **213**, 2621–2633 (2016).
- Finck, R. et al. Normalization of mass cytometry data with bead standards. *Cytom. Part A* **83**, 483–494 (2013).
- McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
- Van Gassen, S. et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytom. Part A* **87**, 636–645 (2015).
- Ingelfinger, F. et al. Single-cell profiling of myasthenia gravis identifies a pathogenic T cell signature. *Acta Neuropathol.* **141**, 901–915 (2021).
- Weber, L. M., Nowicka, M., Sonesson, C. & Robinson, M. D. difcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. *Commun. Biol.* **2**, 183 (2019).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
- Spitzer, M. H. et al. An interactive reference framework for modeling a dynamic immune system. *Science* **349**, 1259–1265 (2015).
- Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
- García-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).
- Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
- Bates, T. C., Maes, H. & Neale, M. C. Umx: twin and path-based structural equation modeling in R. *Twin Res. Hum. Genet.* **22**, 27–41 (2019).
- Verweij, K. J. H., Mosing, M. A., Zietsch, B. P. & Medland, S. E. Estimating heritability from twin studies. *Methods Mol. Biol.* **850**, 151–170 (2012).

Acknowledgements We thank all study participants who took part in this study; L. Robinson from Insight Editing London for manuscript editing; the Core Facility Flow Cytometry unit at the Biomedical Center, Ludwig Maximilian University of Munich, for providing equipment; and the staff from CCGA Kiel for sequencing and transfer of data. This work was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy; ID 390857198), the Gemeinnützige Hertie Stiftung, Bavarian association and National Association of the German MS Society (DMSG), Dr Leopold and Carmen Ellinger Foundation, and the association 'Verein zur Therapieforschung für MS Kranke e.V.'. This project received further funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 882424, the Swiss National Science Foundation (Ambizione PZ00P3_193330 to S.M., and 733 310030_170320, 310030_188450 and CRSII5_183478 to B.B.), as well as the Swiss MS Society Research grant to S.M. and F.I.; F.I. received a PhD fellowship from the Studienstiftung des deutschen Volkes.

Author contributions F.I. designed and performed all of the mass cytometry experiments and performed the bioinformatic analysis of all cohorts. L.A.G., A.F.-H., T.K. and R.H. selected and characterized the cohort of twins with MS. V.K., E.B. and K.M.E. performed the CITE-seq analysis. E.G., S.K. and E.F. helped with the mass cytometry experiments. V.K., E.B., E.F., S.K. and P.Z. helped with the computational analysis. R.F. implemented the statistical model for the variance component analysis of healthy twins. C.P. provided statistical input. C.-A.D. and F.G. provided the InfinityFlow dataset and provided intellectual input. L.A.G., M.K., R.H., S.M., B.S., D.D.F. and P.Z. provided intellectual, scientific and clinical input. B.B., L.A.G., E.B., M.K. and R.H. supervised and funded the study. F.I. and B.B. wrote the manuscript. E.B., M.K., R.H. and L.A.G. edited the manuscript.

Competing interests The authors declare no competing interests.

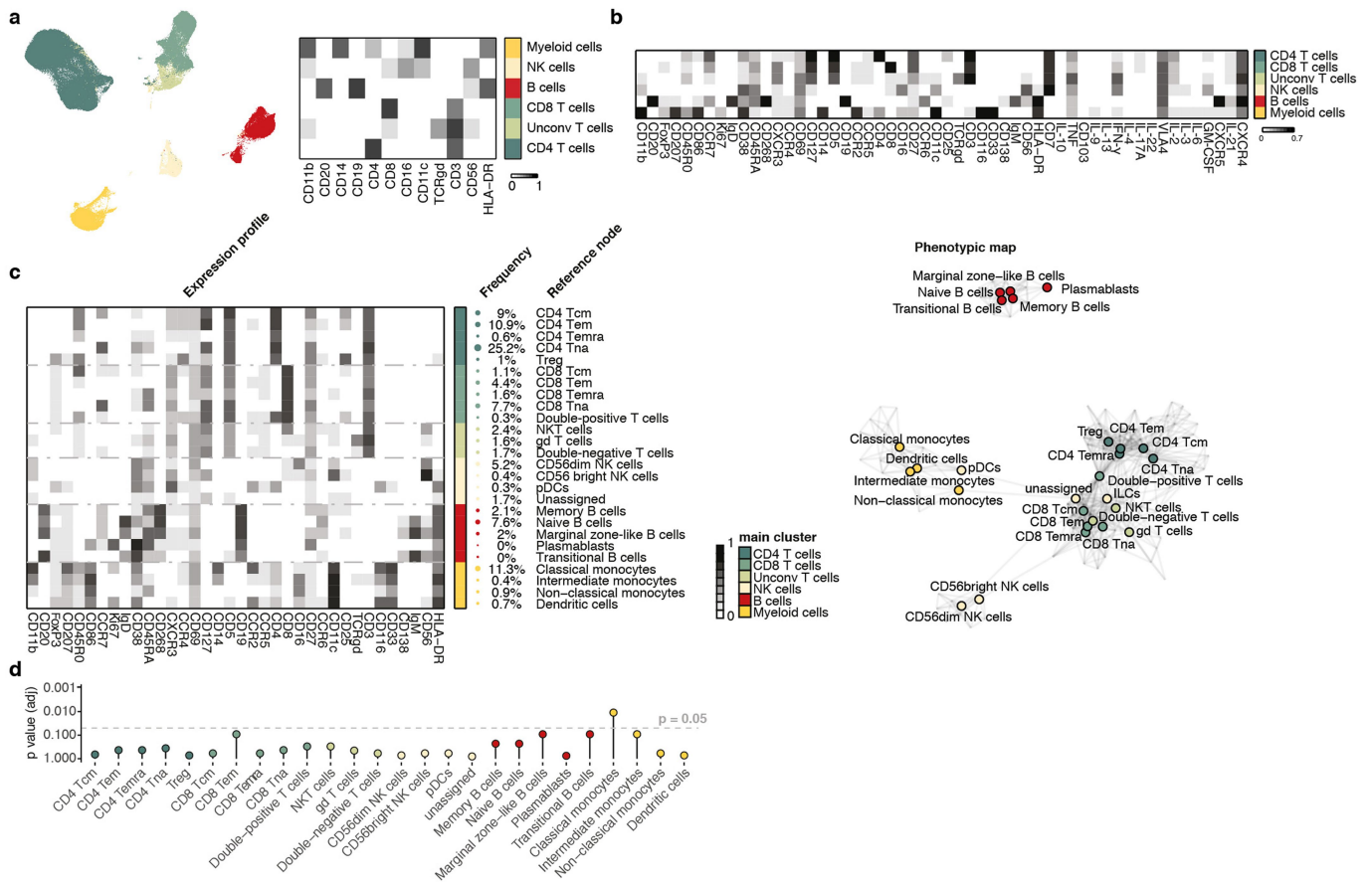
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-04419-4>.

Correspondence and requests for materials should be addressed to Burkhard Becher.

Peer review information Nature thanks Vijay Kuchroo, Lawrence Steinman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

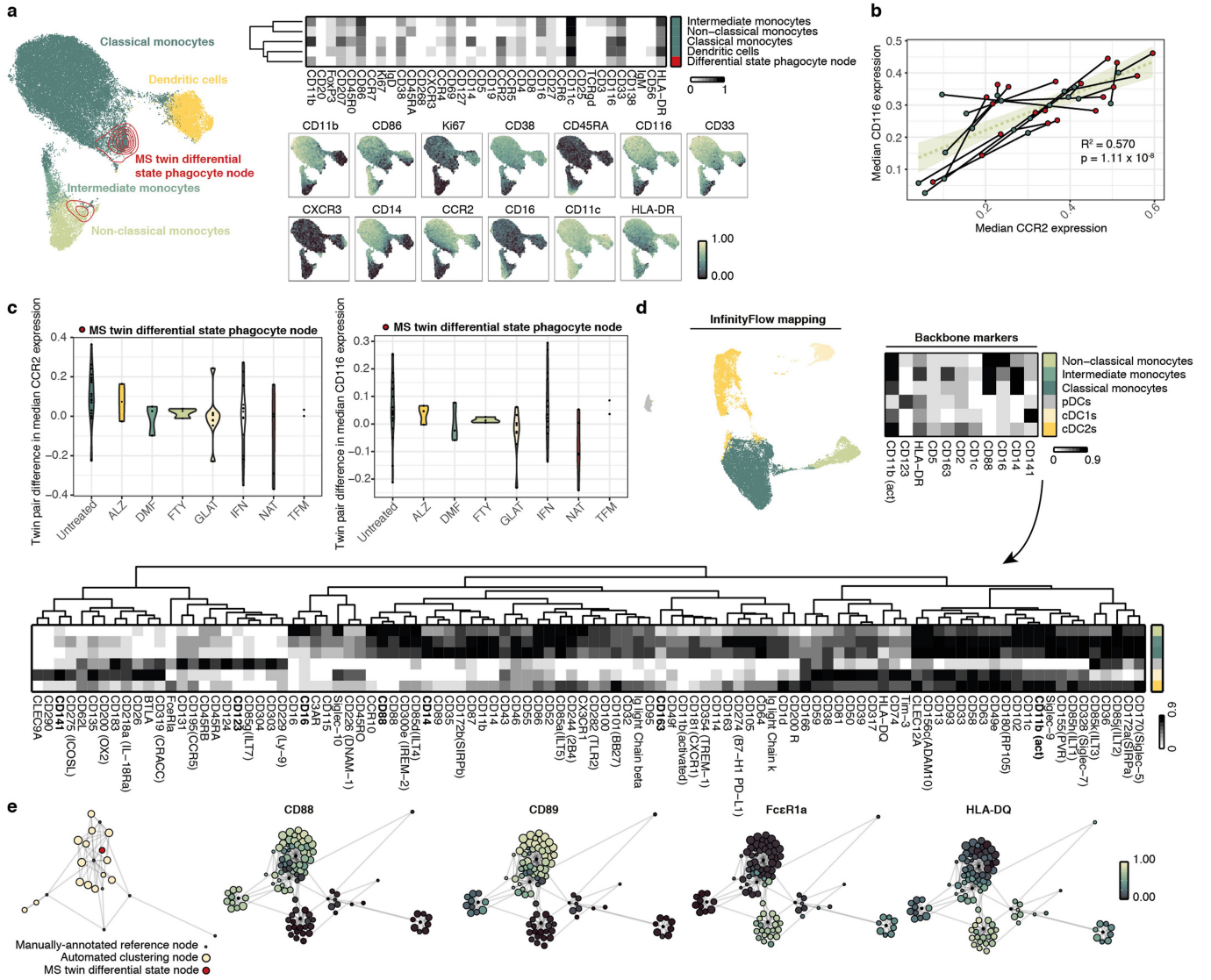
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Changes in canonical immune composition in twin pairs discordant for MS can be attributed to disease-modifying treatment.

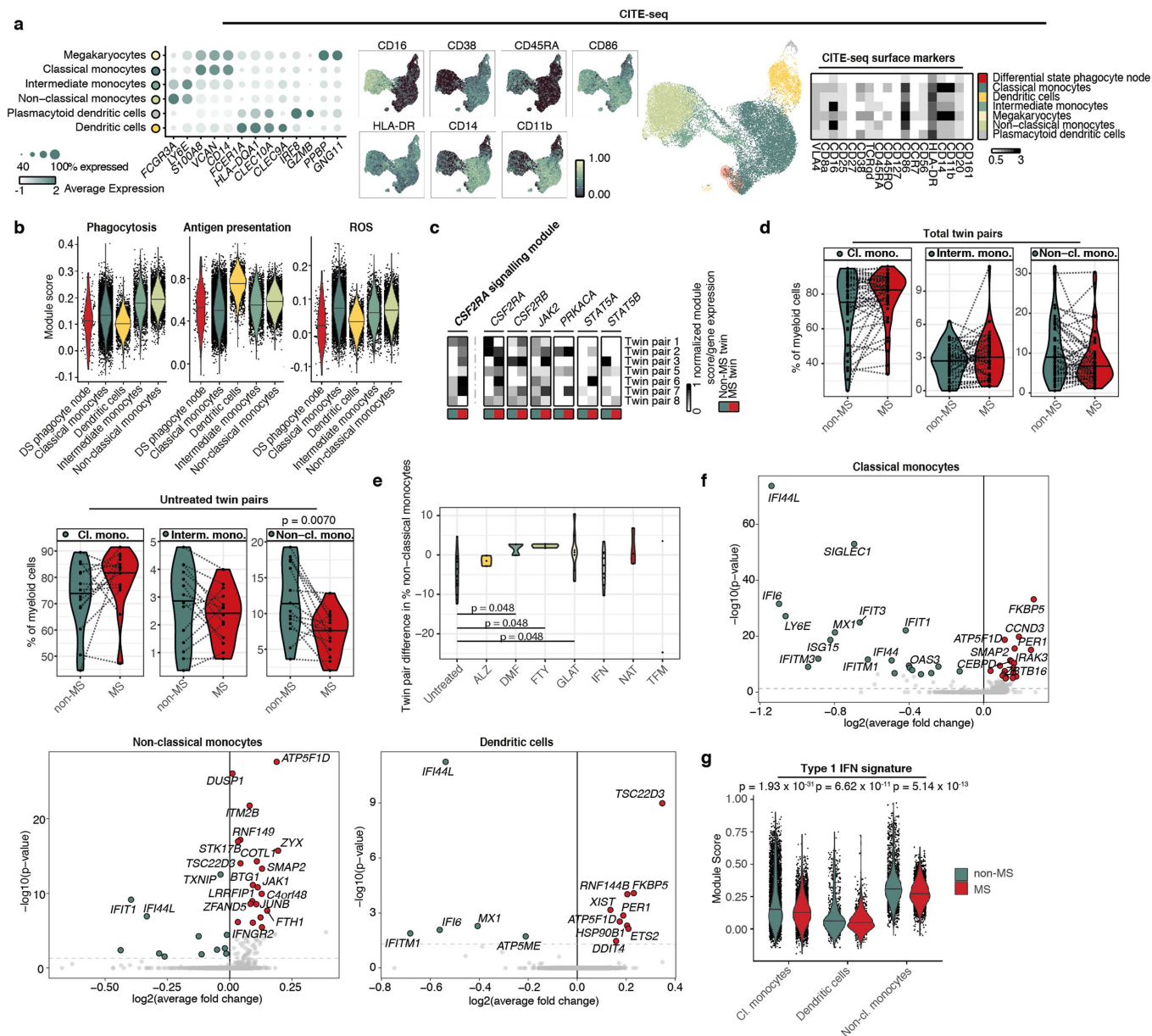
a, UMAP of 1500 cells randomly sampled per patient from the MS twin cohort (left panel). Color code indicates FlowSOM clustering and manual annotation using the lineage markers presented in the heatmap (right panel). **b**, Heatmap showing the combined expression profiles of antigens analyzed by two partially-overlapping mass cytometry panels for the indicated immune populations. **c**, Expression profiles in the CyTOF dataset of the indicated markers across the immune cell populations associated with each reference node, alongside their relative mean population frequencies across all

individuals in the MS twin cohort (n = 114). The FlowSOM-derived clusters were generated by iteratively clustering and manually merging the six main populations described in Extended Data Fig. 1a. **b**, Network representation of immune cell subsets in the manually-annotated reference framework based on phenotypic similarity (right panel). **d**, Lollipop plot showing p-values for the two-sided paired non-parametric Wilcoxon signed-rank test with a false-discovery correction according to the Benjamini-Hochberg approach of immune populations in the reference framework for all MS twins (n = 57) and controls (n = 57).



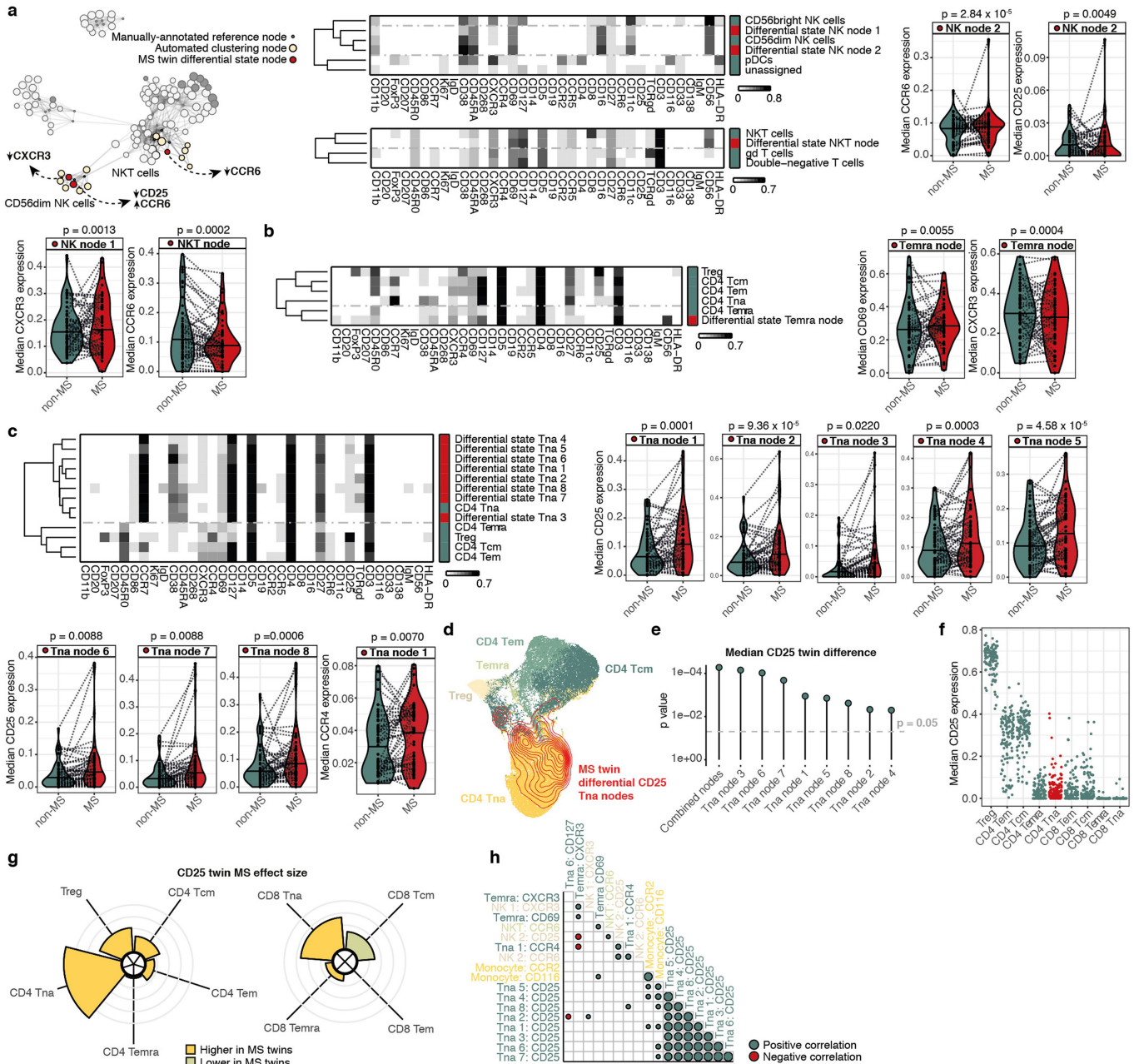
Extended Data Fig. 2 | MS-affected twins show functional perturbations in the monocyte compartment. **a**, UMAP of 1500 randomly sampled myeloid cells per patient with color code indicating nodes of the myeloid reference nodes. Red color indicates density overlay of the differential state phagocyte node that showed increased CCR2 and CD116 expression in MS-affected compared to unaffected twins (left panel). Heatmap (top right panel) displays expression profile of differential state phagocyte node and reference nodes in phenotypic proximity. UMAP with expression overlay of markers present in the myeloid reference nodes (bottom right panel). **b**, 2D lollipop plot showing the median expression of CCR2 and CD116 levels in the differential state phagocyte node from unaffected twins (green) and MS-affected twins (red) that did not receive disease modifying treatment. Twin pairs are indicated by a solid line. Dashed line indicates smoothed conditional mean of the linear regression model with 95% confidence interval in shaded area. **c**, Violin plots showing the inter-twin-pair difference for median CCR2 (left panel) and CD116 (right panel) expression level in cells within the differential state phagocyte node, separated by treatment of the MS-affected twin. ALZ = Alemtuzumab (n = 3); DMF = Dimethyl Fumarate (n = 2); FTY = Fingolimod (n = 4); GLAT = Glatiramer

Acetate (n = 8); IFN = Type 1 Interferons (n = 14); NAT = Natalizumab (n = 4); TFM = Teriflunomide (n = 2). **d**, UMAP of the myeloid cell compartment within healthy PBMCs outside the MS twin cohort from InfinityFlow data used to map the differential state phagocyte node of the MS twin dataset (top left panel). Color code indicates FlowSOM clustering and manual merging of the extended myeloid reference nodes by using the backbone markers presented in the heatmap showing the expression values (top right panel). The heatmap in the bottom panel displays the 100 most differentially expressed markers between the manually-annotated reference nodes. InfinityFlow backbone markers are presented in bold. **e**, Mapping of differentially expressed markers across cell clusters in the MS twin dataset (left panel) on InfinityFlow data of healthy PBMCs (right panels) showing the overlay expression of CD88, CD89, FcεR1a, HLA-DQ. Dot size corresponds to population frequency among total leukocytes. Violin plots contain a bold horizontal line depicting the respective group mean. If not indicated, differences between experimental groups were statistically not significant ($p > 5\%$) using a two-sided unpaired nonparametric Mann-Whitney-Wilcoxon test with a false-discovery correction according to the Benjamini-Hochberg approach.



Extended Data Fig. 3 | Transcriptional profiling of phagocytes of MS-affected twins and unaffected twin siblings. **a**, Dot plot showing the expression profile of selected signature markers in CITE-seq data of phagocytes derived from 7 twin pairs discordant for MS (21 043 cells; left panel). Dot size corresponds to the fraction of cells within each cluster expressing the indicated transcript and color indicates average expression. UMAP showing (middle panel) the expression overlay of indicated markers and cell type yielded by clustering. Heatmap showing the expression profiles of the indicated clusters determined by CITE-seq (right panel). **b**, Violin plots (of 20 678 cells) showing the module score for genes involved in phagocytosis, antigen presentation and reactive oxygen species (ROS) production for the indicated phagocyte clusters. **c**, Heatmap showing the module score for genes involved in *CSF2RA* signaling (left box) and normalized expression of individual selected genes involved in *CSF2RA* signaling (right box) for individual MS-affected and unaffected twins in the differential state phagocyte node (162 cells). **d**, Violin plots showing the frequency of manually-annotated reference nodes in proximity to the differential state phagocyte node determined by mass cytometry for all twin pairs ($n = 57$) and untreated twin pairs ($n = 20$).

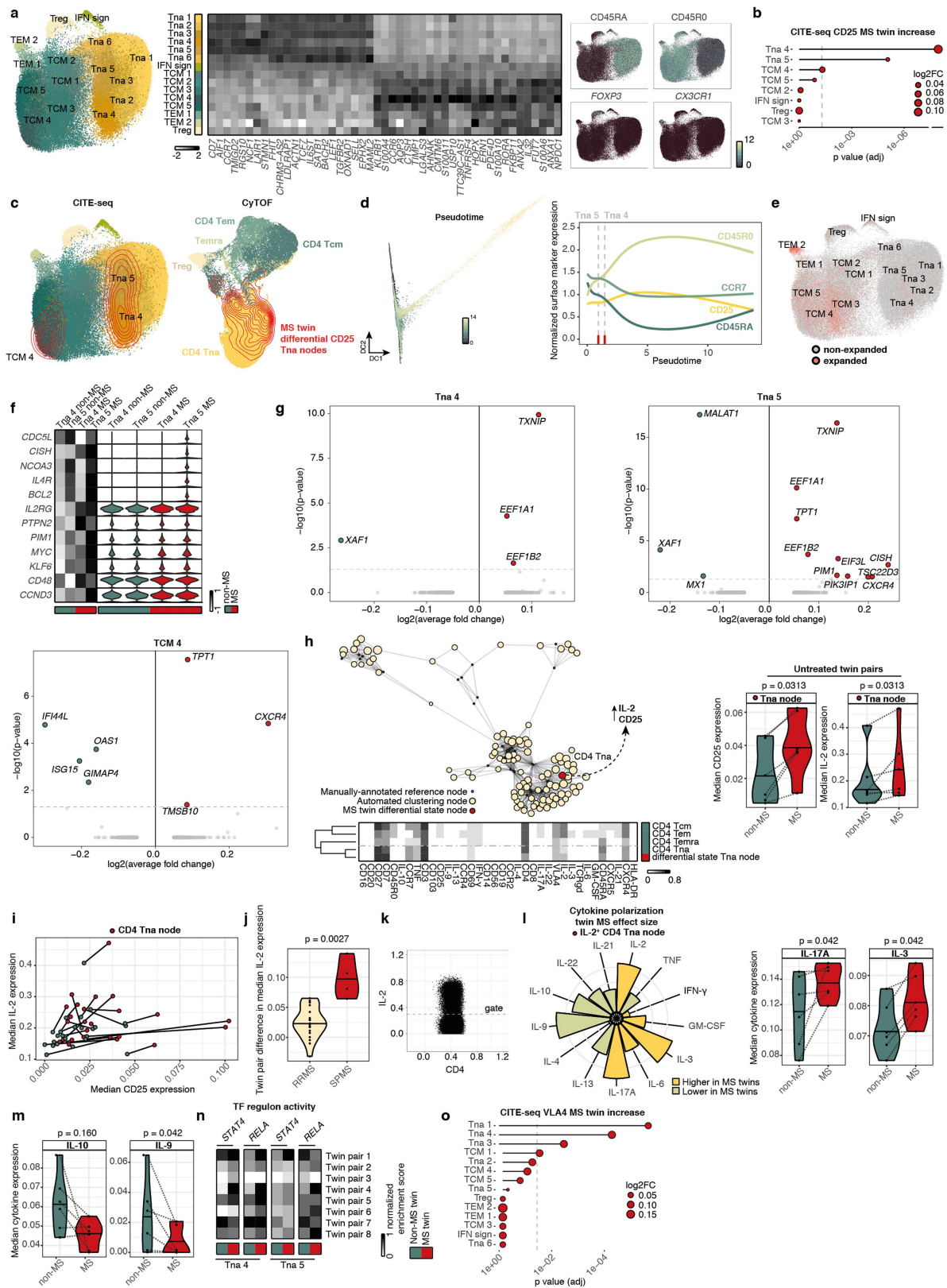
e, Violin plots showing the inter-twin-pair difference in non-classical monocyte frequency, separated by treatment of the MS-affected twin. **f**, Volcano plots showing the differentially expressed genes for the indicated clusters (a total of 18 024 cells) of MS-affected twins compared to unaffected twin siblings. The top 20 significant genes were annotated. Red color indicates increased and green color decreased gene expression in MS twins. **g**, Violin plots of CITE-seq data (a total of 18 024 cells) showing the module score of individual cells for type 1 IFN signature in clusters for which significant differences between MS-affected and unaffected twins were identified. Violin plots contain a bold horizontal line depicting the respective group mean. If not indicated, differences between experimental groups were statistically not significant ($p > 5\%$) using a two-sided paired non-parametric Wilcoxon signed-rank test (**d**) or two-sided unpaired non-parametric Mann-Whitney-Wilcoxon test (**e**; both applying a false-discovery correction according to the Benjamini-Hochberg approach) or using a logistic regression model with the twinning as a latent variable and comparing each model to a null model with a likelihood ratio test and applying a Bonferroni correction (**f**, **g**).



Article

Extended Data Fig. 4 | MS-affected twins are characterized by a transitional Th cell node displaying increased CD25 expression. **a**, Network visualization of unsupervised FlowSOM clusters in the unconventional T cell and NK cell compartment (left panel) yielded by *diffcyt* analysis (beige; automated clustering nodes); highlighted are MS twin differential state nodes for which features appeared significantly different in both filters (red) and nodes of the manually-annotated reference framework (green). Dot size corresponds to population frequency among total leukocytes. Heatmap showing the expression profile of MS twin differential state nodes for immune features present in the unconventional T cell and NK cell compartment and manually-annotated reference nodes in phenotypic proximity (middle panel). Violin plots showing the indicated median marker expression level in cells within the differential state NK and NKT cell nodes for all twin pairs enrolled in the study (n = 57; right panel). **b**, Heatmap showing the expression profile of MS twin differential state Temra node and manually-annotated reference nodes in phenotypic proximity (left panel). Violin plots showing the median CD69 and CXCR3 expression level in cells within the differential state Temra nodes for all twin pairs enrolled in the study (n = 57; right panel). **c**, Heatmap showing the expression profile of MS twin differential state nodes for immune features present in the naïve Th cell compartment and manually-annotated reference nodes in phenotypic proximity (left panel). Violin plots showing the median CD25 and CCR4 expression level in cells within the differential state naïve Th cell nodes for all twin pairs (right panel). **d**, UMAP of 1500 randomly sampled Th

cells per patient with color code indicates FlowSOM clustering and manual merging of the Th cell reference nodes presented in Extended Data Fig. 1c. Red color indicates density overlay of the differential state naïve Th cell nodes that showed significantly increased CD25 expression between MS-affected twins and unaffected twin siblings. **e**, Lollipop plot showing p-values for the median CD25 expression levels across all differential state naïve Th cell nodes individually and combined for all MS-affected twins (n = 57) and unaffected twin siblings (n = 57; left panel). **f**, Dot plot showing the median CD25 expression level among cells within all T cell reference nodes for all patients enrolled in the study (n = 114). **g**, Effect size for the twin-pair difference in median CD25 expression across cells from the manually-annotated reference Th cell (left panel) or cytotoxic T cell nodes (right panel) for all twin pairs (n = 57). **h**, Correlogram showing the correlations between significant immune features for all individuals (n = 114). Dot size indicates the magnitude of correlation coefficient, green dots indicate a positive correlation and red dots a negative correlation. Violin plots contain a bold horizontal line depicting the respective group mean and dashed line indicates twinship. If not indicated, the differences between experimental groups were statistically not significant (p > 5%) using a two-sided paired non-parametric Wilcoxon signed-rank test (**e**, **g**) or using the moderated *limma-trend* method implemented in *diffcyt* applying a false discovery correction according to the Benjamini-Hochberg approach (**a–c**).

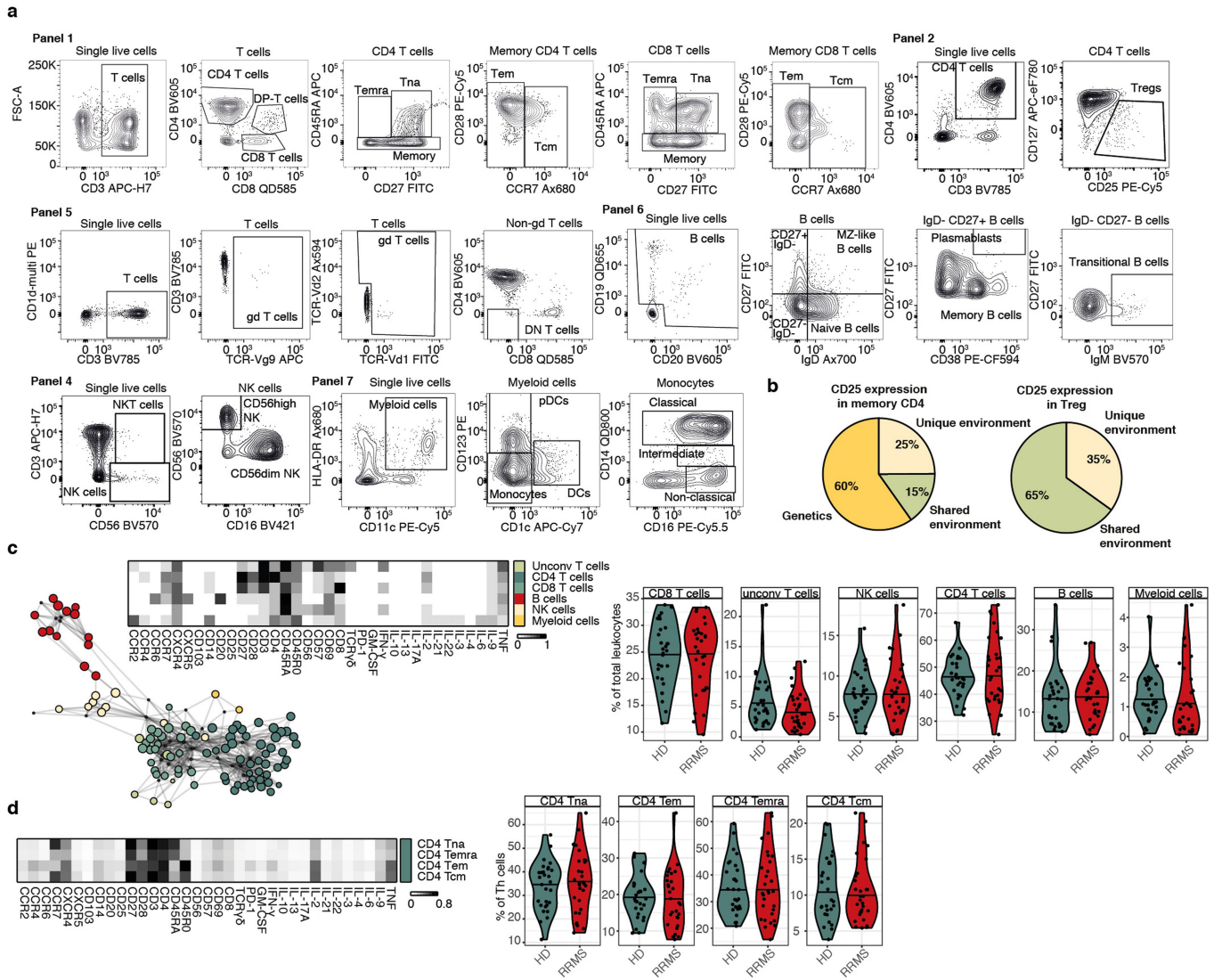


Extended Data Fig. 5 | See next page for caption.

Article

Extended Data Fig. 5 | Transitional Th cells from MS-affected twins respond with increased IL-2 expression upon ex vivo reactivation. **a**, UMAP showing CITE-seq data of Th cells (78 531 cells) derived from 8 twin pairs discordant for MS (left panel). Color coding indicates cell type yielded by clustering using the transcriptome. Heatmap showing the expression profile of the top differentially expressed genes across the different clusters (middle panel). UMAP showing the expression overlay for surface CD45RA and CD45RO and *FOXP3* and *CX3CR1* transcripts. **b**, Lollipop plot showing p-values for increase in CD25 surface epitope expression between MS-affected twins and unaffected twins for the indicated clusters determined by CITE-seq (45 146 cells). **c**, UMAP of CITE-seq and CyTOF data with an overlay showing the transitional Th cell nodes that demonstrated a differential CD25 expression in MS-affected versus unaffected twins. **d**, Diffusion map Th cells (78 531 cells) with pseudotime overlay (left panel). Line plot (right panel) showing the indicated surface markers expression across the computed pseudotime. Horizontal lines indicate mean pseudotime for Tna 4 and Tna 5 respectively. **e**, UMAP of CITE-seq data (78 531 cells) showing expanded cells (orange; > 2 TCR sequences in same patient) and non-expanded cells (grey). **f**, Heatmap and violin plots of selected genes within the IL2RA signaling pathway in Tna 4 and Tna 5 (11 343 cells) for MS-affected and unaffected twins. **g**, Volcano plots showing the differentially expressed genes of the indicated clusters (16 439 total cells) of MS-affected twins compared to unaffected twin siblings. The top 20 significant genes were annotated. Red color indicates increased and green color decreased gene expression in MS-affected twins. **h**, Network visualization (top left panel) of unsupervised FlowSOM clusters (beige); highlighted are MS twin differential nodes for which CD25 and IL-2 expression appeared significantly (red) and nodes of the manually-annotated reference framework (green). Dot size corresponds to population frequency among total leukocytes. Heatmap (bottom panel) displays marker expression profile of differential state naïve Th cell node and manually-annotated reference nodes in phenotypic proximity.

Violin plots (right panel) showing the median CD25 and IL-2 expression levels in cells within the differential state naïve Th cell nodes in untreated twin pairs ($n = 6$). **i**, 2D lollipop plot showing the median expression of CD25 and IL-2 for unaffected twin siblings (green) and MS-affected twin siblings (red; $n = 25$). Twin pairs are indicated by solid line. **j**, Violin plots showing the inter-twin-pair difference in median IL-2 expression level in the differential state naïve Th cell node, separated according to disease stage of the MS-affected twin. RRMS = relapsing-remitting MS ($n = 21$); SPMS = secondary progressive MS ($n = 4$). **k**, Biaxial plot showing the IL-2 and CD4 expression levels for cells in the differential state naïve Th cell node. The dashed line indicates the gate to determine IL-2-positive cells. **l, m**, Effect size for the twin-pair difference of indicated cytokines in IL-2 expressing cells in the MS twin differential state node for untreated twin pairs ($n = 6$; **l**, left panel). Violin plots showing the median IL-17A and IL-3 (**l**, right panel) or IL-10 and IL-9 (**m**) expression level in IL-2-expressing cells within the differential state transitional Th cell node for MS-unaffected twins and untreated MS-affected twins. **n**, Heatmap displaying the transcription factor regulon activity for *STAT4* and *RELA* in Tna 4 and Tna 5 (11 343 cells) determined by CITE-seq for individual twins. **o**, Lollipop plot showing p-values for increase in VLA4 surface epitope expression between MS-affected and unaffected twins for the indicated clusters (75 856 cells) determined by CITE-seq. Violin plots contain a bold horizontal line depicting the respective group mean and dashed line indicates twinship. If not indicated, the differences between experimental groups were statistically not significant ($p > 5\%$) using a two-sided unpaired nonparametric Mann-Whitney-Wilcoxon test (**j**) or a two-sided paired non-parametric Wilcoxon signed-rank test (**h, l, m**) with false discovery correction according to the Benjamini-Hochberg approach or a logistic regression model using the twinship as a latent variable and comparing each model to a null model with a likelihood ratio test and applying a Bonferroni correction (**b, g, o**).



Extended Data Fig. 6 | Monozygotic and dizygotic non-MS twin pairs facilitate dissection of genetic and environmental influences on immune composition. **a**, Biaxial plots displaying the gating strategy for cells from monozygotic and dizygotic twin pairs of the non-MS twin cohort in order to obtain comparable cell populations as present in the reference framework of the MS twin cohort. **b**, Pie chart displaying the variance components for CD25 expression in memory Th cells (left panel) and regulatory T cells (Treg; right panel) using the structural equation model and the median CD25 expression in the non-MS twin cohort. **c**, Heatmap (top left panel) displays expression profile of canonical subsets detected in the validation cohort. Network visualization of unsupervised FlowSOM nodes of the validation cohort and mapping onto the manually-annotated reference framework (grey dots; bottom left panel). Annotation provided in the publicly available mass cytometry dataset was used

to match the main clusters in the reference nodes and indicated by the respective color. Dot size corresponds to population frequency among total leukocytes. Violin plots (right panel) showing the frequency of major canonical immune subsets in the validation cohort for HD ($n = 29$) and RRMS patients ($n = 30$). **d**, Violin plots showing the frequency of annotated Th cell nodes in the validation cohort for HD and RRMS patients (right panel). Heatmap (left panel) displays the expression profile for the indicated Th cell populations. Violin plots contain a bold horizontal line depicting the respective group mean. If not indicated, differences between experimental groups were statistically not significant ($p > 5\%$) using a two-sided unpaired nonparametric Mann-Whitney-Wilcoxon test with a false discovery correction according to the Benjamini-Hochberg approach.

Article

Extended Data Table 1 | Clinical and demographic baseline characteristics of all study cohorts

Variable	MS twin cohort CyTOF		MS twin cohort CITE-seq	Healthy twin cohort		non-twin MS cohort	
	Combined MS twins	Untreated MS twins		Healthy monozygotic twins	Healthy dizygotic twins	MS patients	Healthy controls
Number	57	20	8	21	22	30	29
Sex (f/m)	43/14	15/5	5/3	unk	unk	21/9	14/15
Age (mean, range)	43.7 (20-67)	48 (33-67)	39.1 (23-56)	51.6 (44-59)	48.1 (41-54)	35.8 (17-55)	31.2 (20-55)
Disease course							
RRMS	44	11	6	NA	NA	30	NA
SPMS	11	7	2	NA	NA	0	NA
PPMS	2	2	0	NA	NA	0	NA
Disease Duration in years (mean, range)	12.6 (1-41)	16 (1-41)	11.5 (1-27)	NA	NA	1.6 (0-23)	NA
EDSS	3.0 (0-9.5)	4.2 (1-9.5)	2.5 (0-7.5)	NA	NA	1.7 (0-6.5)	NA
Disease modifying treatment							
none	20	20	7	NA	NA	30	NA
IFN	14	0	0	NA	NA	0	NA
GLAT	8	0	0	NA	NA	0	NA
NAT	4	0	0	NA	NA	0	NA
FTY	4	0	0	NA	NA	0	NA
ALZ	3	0	0	NA	NA	0	NA
TFM	2	0	0	NA	NA	0	NA
DMF	2	0	1	NA	NA	0	NA

Overview of baseline characteristics of the groups and cohorts analyzed in the present study.

MS = Multiple Sclerosis; ALZ = Alemtuzumab; DMF = Dimethyl Fumarate; FTY = Fingolimod; GLAT = Glatiramer Acetate; IFN = beta Interferons; NAT = Natalizumab; TFM = Teriflunomide; f = female; m = male; RRMS = relapsing-remitting MS; SPMS = secondary progressive MS; PPMS = primary progressive MS; NA = not applicable; unk = unknown.

Extended Data Table 2 | Overview of results for data-driven analysis of mass cytometry data using diffcyt

Cluster_id	Marker_id	p value untreated twins	p value (adjusted) untreated twins	p value total twins	p value (adjusted) total twins	Label
15	CCR4	1,59E-04	1,41E-02	3,69E-04	6,98E-03	CD4 Tna
15	CD25	4,57E-05	8,99E-03	1,27E-06	1,18E-04	CD4 Tna
16	CD25	3,98E-05	8,99E-03	8,70E-07	9,36E-05	CD4 Tna
33	CD69	4,49E-04	2,52E-02	2,65E-04	5,53E-03	CD4 Temra
33	CXCR3	2,33E-05	8,15E-03	6,30E-06	4,20E-04	CD4 Temra
37	CD25	1,45E-05	6,76E-03	2,11E-03	2,20E-02	CD4 Tna
46	CD25	2,22E-04	1,62E-02	4,31E-06	3,18E-04	CD4 Tna
5	CD25	2,31E-04	1,62E-02	2,95E-07	4,58E-05	CD4 Tna
6	CD127	8,10E-04	4,20E-02	2,66E-05	1,16E-03	CD4 Tna
6	CD25	1,61E-04	1,41E-02	5,55E-04	8,83E-03	CD4 Tna
63	CCR6	1,05E-04	1,17E-02	2,06E-06	1,73E-04	NKT cells
75	CXCR3	1,08E-04	1,17E-02	3,26E-05	1,34E-03	CD56dim NK cells
77	CCR2	5,14E-05	8,99E-03	1,62E-03	1,79E-02	Classical monocytes
77	CD116	6,38E-05	9,92E-03	2,55E-04	5,41E-03	Classical monocytes
8	CD25	1,23E-04	1,23E-02	5,62E-04	8,84E-03	CD4 Tna
9	CD25	2,67E-06	1,87E-03	9,48E-06	5,86E-04	CD4 Tna
95	CCR6	1,85E-04	1,53E-02	1,42E-07	2,84E-05	CD56dim NK cells
95	CD25	4,23E-04	2,46E-02	2,20E-04	4,90E-03	CD56dim NK cells

Shown are cluster-marker combinations that were statistically significant ($p_{adj} < 0.05$) in two comparisons: 1) comparing all MS-affected twins to their unaffected twin siblings and 2) comparing only untreated MS-affected twins to unaffected siblings using the moderated *limma-trend* method implemented in *diffcyt* with a false-discovery correction according to the Benjamini-Hochberg approach.

Tna = naive T cells; Temra = effector memory T cells re-expressing CD45RA.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Fluidigm CyTOF was used to facilitate mass cytometry acquisition.

Data analysis R version 3.6.1 and 4.0.1 was used in RStudio 1.3.959 and Visual Studio Code 1.63 to analyze all data presented. FlowJo 10.6 was used to preprocess the data. Cell Ranger 6.1 was used to process raw sequencing reads. For downstream analysis the following packages were used as indicated in the method section of the manuscript: umap (v0.2.6), FlowSOM (v1.17.4), ConsensusClusterPlus (v1.52.0), diffcyt (v1.10.0), grappolo (v0.5.1), vite (v0.4.10), igraph (v1.2.7), ggraph (v2.0.5), pheatmap (v1.0.12), Hmisc (v4.4.0), corrplot (v0.84), ggplot2 (v3.3.5), Seurat (v4.0.3), DoRothEA (v1.2.2), Monocle 3 (v0.2.3.0), SCANPY (1.6.0), mgcv (v1.8-31), umx (v4.10.10), stats (v4.0.1), rstatix (v0.7.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw mass cytometry data can be accessed at doi:10.17632/fzs5ph5p8s.1. CITE-seq data is available at doi:10.17632/278fy5m2yj.1. Code used for the CITE-seq analysis is accessible at https://github.com/beltranLab/twin_study_Nature_2021. Code used for the mass cytometry and variance component analysis is presented at <https://github.com/florianingelfinger/MStwins>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The MS TWIN STUDY has been established in 2012 and has currently enrolled 87 monozygotic twin pairs with discordance for MS. Samples for the presented MS twin cohort analysis have been selected from the MS TWIN STUDY with respect to availability of sufficient biomaterial (PBMCs samples) until May 2018 (CyTOF) and May 2020 (additional samples for CITE-seq).
Data exclusions	Samples containing less than 100 cells and the data of the corresponding co-twin were excluded in mass cytometry analysis. For CITE-seq only samples with data for both the MS-affected and unaffected sibling were used for downstream analysis. These criteria have been pre-established to ensure a reliable quantification of immune cell frequencies and median marker expressions.
Replication	Validation of the finding in the MS twin cohort have been carried out using CITE-seq of additional twin pairs in the MS twin cohort and using publicly available mass cytometry data of a cross-sectional non-twin MS cohort. Mass cytometry was acquired in two independent runs and was performed once due to limited sample availability of this precious study cohort. CITE-seq analysis was performed once.
Randomization	Twin samples were not allocated to experimental groups as this was an observational study with pre-established clinical conditions (MS, non-MS). However, randomization was carried out to balance the two individual runs of the mass cytometry analysis. For twin pairs were equally randomized and the respective co-twin was analyzed in the same batch as the MS-affected twin.
Blinding	Investigators were blinded during the experiments. After data acquisition and preprocessing unblinding was necessary to perform group stratification. However, the unbiased nature of the analysis approach prevented investigator bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	All mass cytometry antibodies used in the study including manufacturer, clone, heavy metal tag and dilutions are listed in Supplementary Table 3. CITE-seq antibodies are listed in Supplementary Table 7.
Validation	Heavy metal conjugated antibodies have been validated by assessing the respective expression pattern in a fully stained sample across cell types in the human peripheral blood and comparing it to the manufacturers technical sheets and cytometry and sequencing data available in the literature. Clones have been preferentially chosen based on previous experience with the respective antibody for flow cytometry.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Inclusion criteria for study participation were met if in one co-twin of monozygotic twins an MS diagnosis according to the revised McDonald criteria was established, whereas the co-twin was clinically healthy. Exclusion criteria were infection as well as treatment with antibiotics or high dose intravenous glucocorticosteroids within three months prior to sampling. Monozygotic twin pairs clinically discordant for MS (n=61) visited the outpatient department at the Institute of Clinical
----------------------------	---

Neuroimmunology at the LMU Klinikum Munich for a detailed interview, neurological examination, blood sampling and MRI investigations (in a proportion of twins only). To confirm MS diagnosis, medical records including MRI scans were obtained and reviewed (Extended Data Table 1, Supplementary Tables 1-2). 43 female and 14 male twin pairs were analyzed by mass cytometry and age ranged between 20 and 67 years (mean = 43.7 years). For CITE-seq analysis 5 female and 3 male twin pairs at an age between 23 and 56 years (mean = 39.1 years) were analyzed.

Recruitment

The MS twin cohort is part of the MS TWIN STUDY and presents a cohort of monozygotic twins with discordance for MS and is located at the Institute of Clinical Neuroimmunology at the LMU Klinikum Munich, Germany. Recruitment started in May 2012 and is still ongoing, samples used in the present study were collected up to May 2018. Information regarding the MS TWIN STUDY is distributed via an internet appeal to people with MS (such as webpages or magazines of national and regional MS societies, support groups) or to MS neurologists via publications and talks from the PIs from the MS TWIN STUDY. Contact with interested participants is based on a voluntary approach by the twins or a referring neurologist. Recruitment is planned after detailed information is provided during a detailed telephone interview. Potential biases of a study based on a voluntary appeal in contrast to a registry based approach are as follows: Due to limitations to take the trouble to travel to the study site our study cohort might have a larger fraction of younger twins with less severe disease severity whereas people with higher disease severity or older age refuse to take the trouble to travel to the study centre. In the same line a potential bias reflects the outreach of an internet appeal to younger people. In addition, the appeal for a voluntary participation in a twin study to unravel the triggers for MS might attract participants with more than one family member affected with MS and might introduce a higher familial risk in the study cohort than in the general MS population. Another limitation is the heterogeneity of the cohort regarding age, disease course and varying treatments. Disease modifying treatment of the MS affected twin within a twin pair presents the most important bias and this limitation cannot be overcome since most people with MS receive disease modifying treatment and the repertoire of effective treatments has been ever expanding. But as explained in detail these pairs have been excluded for the relevant analyses. However, taking these potential selection biases into account we do not assume that these impact our study results on the immune perturbation in MS.

Ethics oversight

The MS TWIN STUDY was approved by the local ethics committees of the Ludwig-Maximilians-University of Munich (ethics approval project number 267-13). All participants gave written informed consent, according to the principles of the Declaration of Helsinki.

Note that full information on the approval of the study protocol must also be provided in the manuscript.