

Altered human oligodendrocyte heterogeneity in multiple sclerosis

Sarah Jäkel^{1,5}, Eneritz Agirre^{2,5}, Ana Mendenha Falcão², David van Bruggen², Ka Wai Lee², Irene Knuesel³, Dheeraj Malhotra^{3,6}, Charles ffrench-Constant^{1,6*}, Anna Williams^{1,6*} & Gonçalo Castelo-Branco^{2,4,6*}

Oligodendrocyte pathology is increasingly implicated in neurodegenerative diseases as oligodendrocytes both myelinate and provide metabolic support to axons. In multiple sclerosis (MS), demyelination in the central nervous system thus leads to neurodegeneration, but the severity of MS between patients is very variable. Disability does not correlate well with the extent of demyelination¹, which suggests that other factors contribute to this variability. One such factor may be oligodendrocyte heterogeneity. Not all oligodendrocytes are the same—those from the mouse spinal cord inherently produce longer myelin sheaths than those from the cortex², and single-cell analysis of the mouse central nervous system identified further differences^{3,4}. However, the extent of human oligodendrocyte heterogeneity and its possible contribution to MS pathology remain unknown. Here we performed single-nucleus RNA sequencing from white matter areas of post-mortem human brain from patients with MS and from unaffected controls. We identified subclusters of oligodendroglia in control human white matter, some with similarities to mouse, and defined new markers for these cell states. Notably, some subclusters were underrepresented in MS tissue, whereas others were more prevalent. These differences in mature oligodendrocyte subclusters may indicate different functional states of oligodendrocytes in MS lesions. We found similar changes in normal-appearing white matter, showing that MS is a more diffuse disease than its focal demyelination suggests. Our findings of an altered oligodendroglial heterogeneity in MS may be important for understanding disease progression and developing therapeutic approaches.

We performed single-nucleus RNA sequencing (snRNA-seq) from the white matter of post-mortem tissue of five human controls without neurological disease and four individuals with progressive MS (Supplementary Table 1) using the 10x Genomics pipeline⁵ (Extended Data Fig. 1a). We isolated nuclei from different white matter areas within the same MS tissue block or patient, including normal-appearing white matter (NAWM), active, chronic active, chronic inactive and remyelinated lesions (Extended Data Fig. 1b) as defined by neuropathology⁶. After quality control, we obtained 17,799 nuclei, with a mean of 1,096 genes per nucleus, and a mean of 1,795 unique molecular identifiers (UMIs) per nucleus (Extended Data Fig. 1e, f and Supplementary Table 2).

We performed canonical correlation analysis (CCA) in the combined dataset of control individuals and patients with MS, to minimize sample individual variability batch effects, and clustering with Seurat⁷ (Extended Data Fig. 3). We identified five subclusters of neurons, seven of oligodendrocytes and additional clusters for oligodendrocyte progenitor cells (OPCs), committed oligodendrocyte precursors (COPs), astrocytes, vascular smooth muscle cells, pericytes, endothelial cells and immune cells (Fig. 1a, Extended Data Fig. 1g and Supplementary Table 3). We found unique or enriched RNA markers for the individual subclusters within the oligodendrocyte lineage (Fig. 1d, Extended Data Fig. 1c and Supplementary Table 4): *PDGFRA*, *BCAN* and *SOX6*

for OPCs, *APOE* and *CD74* for immune oligodendroglia (ImOLG; see below), *CDH20* and *RBFox1* for Oligo1, *LURAP1L-AS1* and *CDH19* for Oligo2, *KLK6* and *GJB1* for Oligo5, and *OPALIN* and *LINC00844* for Oligo6. We confirmed the presence of *OLIG2* and the absence of *NOGO* (also known as *RTN4*) in human OPCs, as used for their identification by neuropathologists⁸ (Extended Data Fig. 4a). Immunohistochemistry (IHC) analysis showed that these *OLIG2*⁺*NOGO*⁻ OPCs are also *SOX6*⁺ (Extended Fig. 4b, e). In addition, IHC confirmed co-labelling of *KLK6* (Oligo5) or *OPALIN* (Oligo6) with *OLIG2* (Extended Data Fig. 4c, d) and segregation of Oligo5 and Oligo6 (Fig. 1b and Extended Data Fig. 4f-h) on a different set of donor tissue. Segregation of pairs of subcluster markers for Oligo1, Oligo2 and Oligo5 was also confirmed using duplex in situ hybridization, with less than 10% of oligodendrocytes containing both RNA markers (Fig. 1c). Correlation analysis with oligodendroglia from an experimental autoimmune encephalomyelitis (EAE) mouse model of MS⁴ indicated similarities between mouse and human OPCs (Extended Data Fig. 5 and Supplementary Table 5). Human Oligo1 and Oligo5 correlated with mouse mature oligodendrocytes (MOL) 1 and MOL2, whereas the remaining mature human oligodendrocyte populations were closer to mouse MOL5 and MOL6. Human Oligo3 and ImOLG also presented similarities to mouse OPCs and COPs (Extended Data Fig. 5). Therefore, human white matter has transcriptionally heterogeneous oligodendrocyte states that show some similarities to the adult mouse counterpart.

In accordance with adult mouse brain single-cell RNA sequencing (scRNA-seq)³, we detected very few cells with the hallmarks of newly formed oligodendrocytes (Fig. 1a). Thus, we combined our data from control white matter with previously published adult brain data^{9,10} using CCA analysis followed by clustering with Seurat⁷ (Extended Data Fig. 2a and Supplementary Table 2). We were able to re-identify some of our oligodendrocyte subclusters in these other datasets (Extended Data Fig. 2a). Moreover, we now found that the Oligo6 cluster had hallmarks of an intermediate state between OPCs and mature oligodendrocytes (Extended Data Fig. 2c, d). To confirm Oligo6 as an intermediate oligodendrocyte state, we performed single-cell near-neighbour network embedding (SCN3E) analysis¹¹ to order the identified populations in pseudotime (Fig. 2a). A subset of Oligo6 nuclei connected OPCs and COPs with the remaining mature oligodendrocytes, confirming the intermediate character of this cluster. Oligo1 and Oligo5 clusters, by contrast, represented end states in the SCN3E analysis. Notably, Gene Ontology (GO) analysis indicated that the highest-expressed genes in these clusters were not myelin genes (Extended Data Fig. 6), which suggests that these mature stable oligodendrocytes do not need to maintain strongly active transcriptional machinery for myelination but instead a transcriptional network that reinforces signalling, cell-to-cell adhesion and viability. By contrast, GO analysis indicated that the Oligo3 and Oligo4 clusters represented actively myelinating oligodendrocytes, expressing genes associated with the myelination and membrane assembly pathways (Extended Data Fig. 6).

¹MRC Centre for Regenerative Medicine and MS Society Edinburgh Centre, Edinburgh bioQuarter, University of Edinburgh, Edinburgh, UK. ²Laboratory of Molecular Neurobiology, Department of Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. ³Roche Pharma Research and Early Development (pRED), Roche Innovation Center Basel, Basel, Switzerland. ⁴Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm, Sweden. ⁵These authors contributed equally: Sarah Jäkel, Eneritz Agirre. ⁶These authors jointly supervised the work: Dheeraj Malhotra, Charles ffrench-Constant, Anna Williams, Gonçalo Castelo-Branco. *e-mail: cffc@ed.ac.uk; anna.williams@ed.ac.uk; goncalo.castelo-branco@ki.se

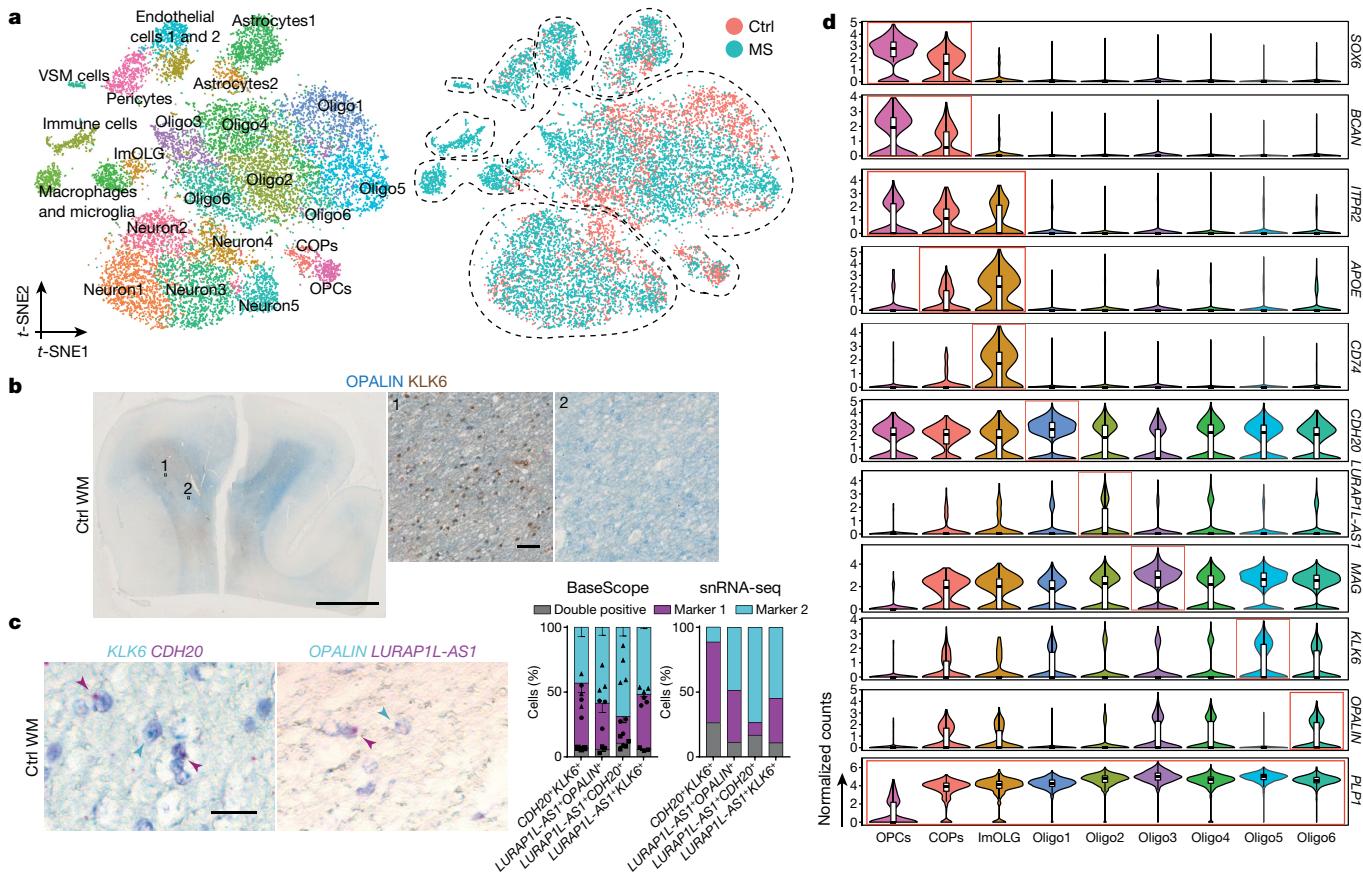


Fig. 1 | snRNA-seq reveals oligodendroglia heterogeneity in the human brain. **a**, t-SNE projection of all recovered cell clusters, sorted by cell population (left) or disease condition (right) ($n = 17,799$ nuclei from 5 control (ctrl) individuals and 4 patients with MS). VSM, vascular smooth muscle. **b**, Combined OPALIN and KLK6 staining of human control white matter (WM). Scale bars, 5 mm and 50 μ m (insets). **c**, Left, double in situ hybridization (BaseScope assay) of human control white matter counterstained with haematoxylin. Right, quantification of the proportion of double-positive oligodendrocytes determined by in situ hybridization (left graph) and the snRNA-seq dataset (right graph). Left graph: $n = 4$ for $LURAP1L-AS1^+$ - $CDH20^+$, $n = 3$ for other combinations. Experiments were performed in three independent batches. Data are mean \pm s.e.m.

We next compared the single-nucleus transcriptional profiles of oligodendroglia from the individuals with MS with the control individuals. CCA analysis (considering all the individual samples as a variable and using the union of the top variable genes from each of the samples) and Seurat clustering led to the identification of all brain cell types, including pericytes, macrophages and other immune cells (Fig. 2b), reflecting immunological infiltration of the central nervous system in MS. The total number of oligodendrocyte nuclei isolated in control and MS samples was within the same range (Fig. 2d). We quantified OLIG1/OLIG2-expressing cells from a different patient cohort and, despite there being fewer cells generally in demyelinated lesions, the percentage of OLIG1/OLIG2 $^+$ cells in lesions did not change in these lesions compared with NAWM and control white matter (Fig. 3a). Further analysis revealed the same oligodendroglial subclusters in the nuclei derived from patients with MS as in control individuals. However, the frequency of nuclei in individual subclusters was markedly different between control and MS samples in three ways (Fig. 2b), which helps to explain previous microarray analyses of human brain tissue that show different oligodendrocyte transcriptional outputs in MS brain at the population level^{12–16}.

First, we observed fewer nuclei from OPCs in all MS lesions and in NAWM (Fig. 2c, d). To verify this reduction in other patients with MS, we quantified OPCs using the specific markers identified above for

Rectangles, circles and triangles denote individual values of double-positive markers, marker 1 and marker 2, respectively. Right graph: single nucleus considered positive when normalized expression > 0 for marker 1, marker 2 or both. $n = 5,902$ for $CDH20^+$ - $KLK6^+$; $n = 2,980$ for $LURAP1L-AS1^+$ - $OPALIN^+$; $n = 5,782$ for $LURAP1L-AS1^+$ - $CDH20^+$; and $n = 3,395$ for $LURAP1L-AS1^+$ - $KLK6^+$. **d**, Violin plots of markers enriched in specific oligodendrocyte subpopulations showing normalized gene expression. $n = 352$ for OPC; $n = 242$ for COP; $n = 207$ for ImOLG; $n = 1,129$ for Oligo1; $n = 1,839$ for Oligo2; $n = 775$ for Oligo3; $n = 1,579$ for Oligo4; $n = 1,167$ for Oligo5; and $n = 1,484$ for Oligo6. Violin plots are centred around the median with interquartile ranges, and the shape represents cell distribution.

human OPCs, BCAN and SOX6, on post-mortem MS tissue from a different patient cohort (Supplementary Table 1). Using both IHC against SOX6 (Fig. 3b) and in situ hybridization against BCAN (Extended Data Fig. 8a), we confirmed a significant reduction in the number of OPCs both in lesions and NAWM, compared to control tissues. This is consistent with previous studies^{17–19} of OPC numbers showing their loss in some MS lesions.

Second, the intermediate Oligo6 cells were highly reduced in MS (Fig. 2b, d). We confirmed this using IHC against OPALIN on MS tissue, both in lesions and in NAWM (Fig. 3c). In addition, we found that OPALIN $^+$ Oligo6 cells were predominantly localized to the junction between the white and grey matter (Extended Data Fig. 4f, g). This widespread decrease of both OPALIN $^+$ cells and OPCs in MS tissue adds evidence to the concept that NAWM is indeed not ‘normal’ but has more global changes that may reflect a propensity to demyelination^{20,21} or a regenerative response.

Third, we identified skewing in the subclusters of mature oligodendroglia between MS and control tissue (Fig. 2c, d and Extended Data Fig. 7e): the Oligo1 cluster was depleted in MS, whereas the Oligo2, Oligo3, Oligo5 and ImOLG clusters were enriched. This skewed distribution remained after deconvolution of our MS samples according to whether they were from NAWM or lesions (Fig. 2c). Using IHC, we confirmed that KLK6 $^+$ Oligo5 cells were not lost in MS lesions

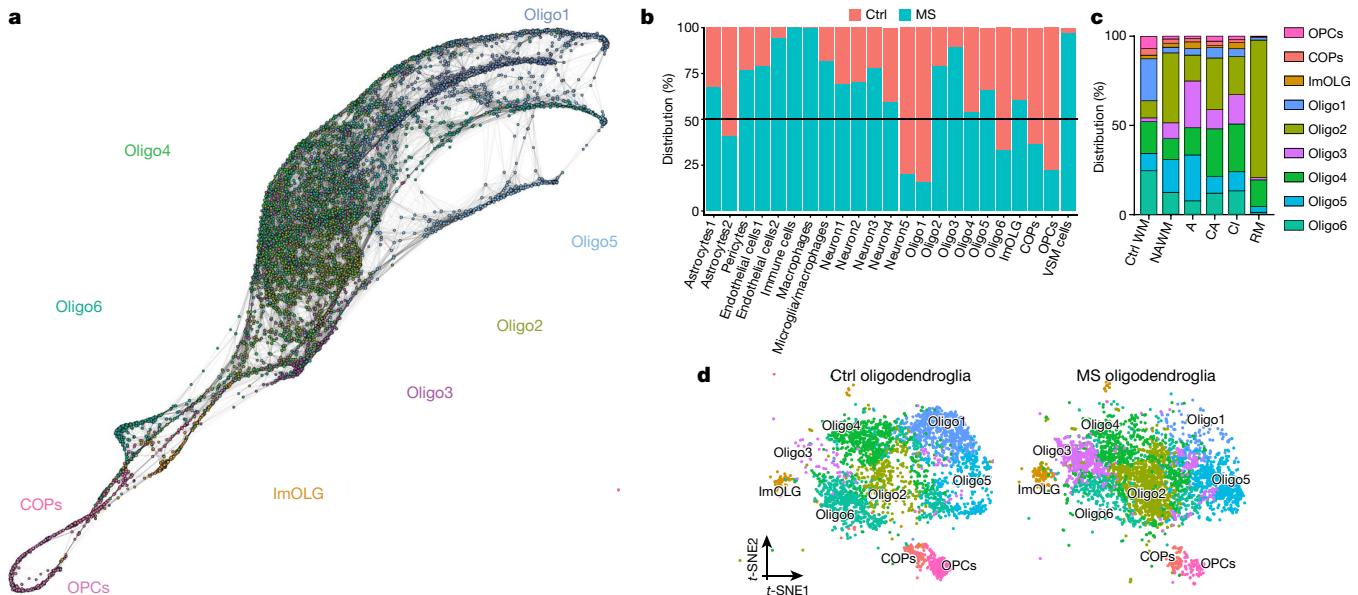


Fig. 2 | Altered oligodendroglia heterogeneity in human MS brain.
a, SCN3E pseudotime analysis of the human oligodendrocyte lineage in control and MS white matter. b, Frequency distribution of all clusters between control (red) and MS (turquoise) nuclei. c, Frequency distribution

of oligodendrocyte clusters between control and different MS lesions. A, active; CA, chronic active; CI, chronic inactive; RM, remyelinated. d, t-SNE projections of oligodendrocyte subclusters in control and MS tissue ($n = 4,037$ control oligodendrocytes; $n = 4,737$ MS oligodendrocytes).

and NAWM (Extended Data Fig. 8b). Although ImOLG expressed canonical oligodendrocyte genes, they were slightly separated from the main oligodendrocyte cloud in the *t*-distributed stochastic neighbour

embedding (*t*-SNE) analysis, were closely associated with microglia and expressed genes such as *CD74*, *HLA-DRA*, *PTPRC* and *C3* (Fig. 3d and Extended Data Fig. 8c). We validated the expression of *CD74* in

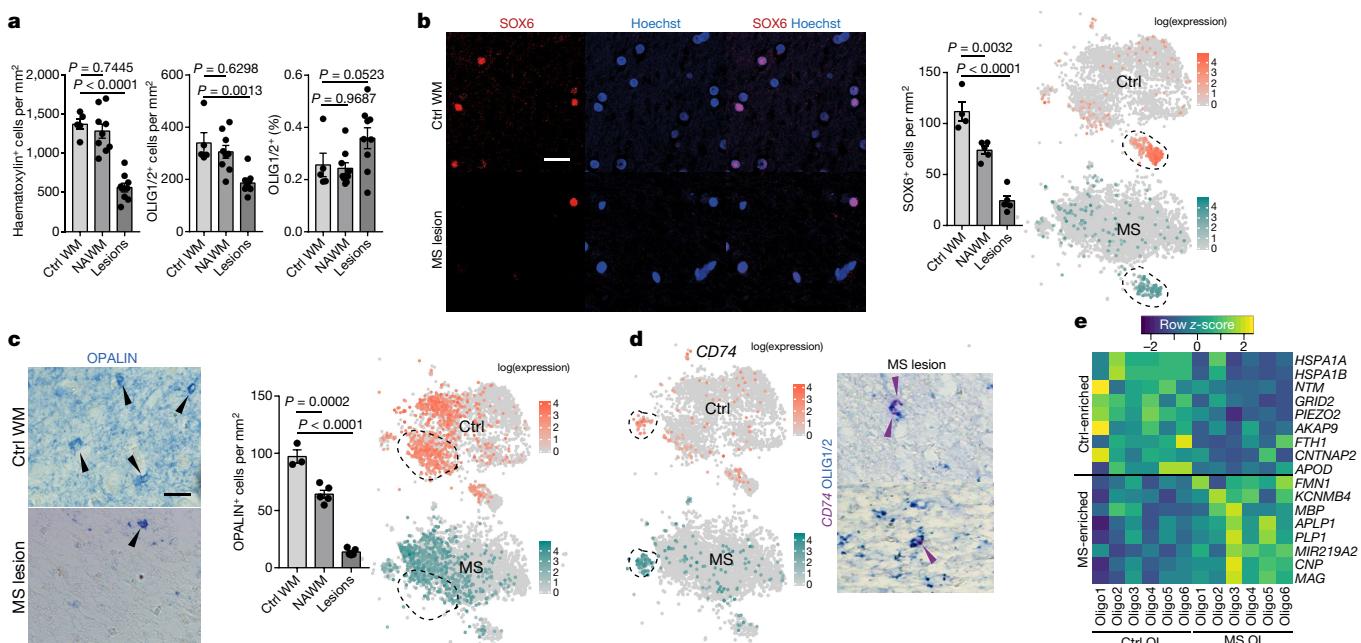


Fig. 3 | Depletion of specific oligodendrocyte subclusters and increased expression of myelination genes in mature oligodendrocytes in human MS brain. a, Total cellular and oligodendrocyte densities in control white matter, NAWM and MS lesions. $n = 5$ control individuals; $n = 9$ MS individuals. b, Left, SOX6-expressing OPCs in control white matter, NAWM and MS lesions. Scale bar, 50 μ m. $n = 4$ control individuals; $n = 5$ MS individuals. Right, *t*-SNE overlay of SOX6 expression in the control and MS snRNA-seq dataset. c, Left, OPALIN-expressing oligodendrocytes in control white matter, NAWM and MS lesions. Scale bar, 50 μ m. $n = 3$ control individuals; $n = 5$ MS individuals. Right, *t*-SNE overlay of OPALIN expression in the control and MS snRNA-seq dataset. d, *CD74* expression

in the control and MS snRNA-seq dataset, and BaseScope *in situ* validation of the presence of *CD74* combined with IHC staining for *OLIG1*/*OLIG2*⁺ oligodendrocytes. $n = 2$ different patients with MS; experiments were performed in two independent batches. e, Heat map representing the average gene expression of a subset of genes, including myelin-related genes, in mature oligodendrocytes in control versus MS samples. For *t*-SNEs and heat map, $n = 4,037$ control oligodendrocytes; $n = 4,737$ MS oligodendrocytes. P values determined by analysis of variance (ANOVA). In a–c, only P values compared with control are shown. Data are mean \pm s.e.m.

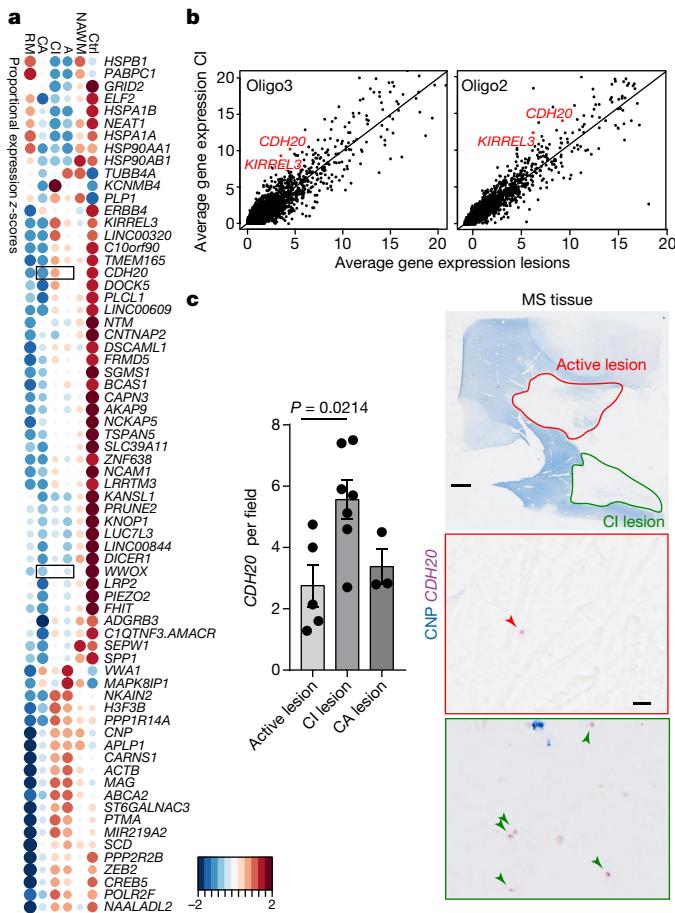


Fig. 4 | Differential gene expression analysis of MS lesions reveals potential specific markers. **a**, Dot plot illustrating the top differentially expressed genes (in terms of percentage of cells expressing these genes per sample) between lesions, NAWM and control tissue; both size and colour indicate z -scores (blue and large denote low scores; red and large denote high scores; and small denotes intermediate scores). Validated genes *CDH20* and *WWOX* are highlighted by squares. **b**, Average gene expression across Oligo2 (left) and Oligo3 (right) subclusters in chronic inactive lesions compared with the average expression in the rest of the lesions. Red denotes examples of genes significantly differentially expressed and upregulated in chronic inactive lesions. Adjusted $P < 0.05$, Bonferroni-corrected Wilcoxon rank sum two-sided test. **c**, BaseScope in situ validation of *CDH20* expression in different lesion types. Demyelinated lesions were detected by the absence of IHC staining for the myelin protein CNP. Scale bars, 2 mm (top panel) and 10 μ m (middle/bottom panel). Data are mean \pm s.e.m. $n = 5$ active lesions, $n = 7$ chronic inactive lesions, and $n = 3$ chronic active lesions derived from $n = 7$ different patients with MS. P values determined by ANOVA (only significant P values are shown).

OLIG1/OLIG2⁺ oligodendrocytes by in situ hybridization (Fig. 3d). GO and SCN3E analysis suggested that this population consists of intermediate oligodendrocytes with an immunological phenotype (Fig. 2a and Extended Data Fig. 6) that we have previously described in mouse experimental autoimmune encephalomyelitis and in human⁴.

In addition, differential gene expression analysis between oligodendrocytes from control individuals and patients with MS indicated that several myelin protein genes were upregulated in mature oligodendrocytes in MS (Fig. 3e, Extended Data Fig. 8h and Supplementary Table 6). We found similar upregulation of myelin genes when comparing control and NAWM (Extended Data Fig. 8h), indicating that, in the context of disease, mature oligodendrocytes might increase transcriptional programs responsible for myelination.

Our dataset has the advantage that we could identify and compare expression of potential biological markers in different lesion types,

NAWM and control white matter, albeit on a limited number of patients. The proportion of cells expressing several genes (for example, *KIRREL3*, *CDH20*, *PLCL1*, *LINC00609*, *FRMD5*, *LRRTM3* and *C1QTNF3-AMACR*) was enriched in control and chronic inactive lesions, but not in other lesion types (Fig. 4a). Other markers were proportionally enriched or depleted in other lesion types, such as *NKAIN2* (reduced proportion in remyelinated lesions) or *WWOX* (reduced proportion in chronic active lesions) (Fig. 4a). Significant differences in expression levels were also observed in specific oligodendrocyte subclusters for *KIRREL3* and *CDH20* (Fig. 4b and Supplementary Table 7). Paradoxically, the average expression (total normalized RNA counts) for some of these genes was lower in control tissues and higher in remyelinated lesions (Extended Data Fig. 9d). This may be due to a lower proportion of cells in remyelinated lesions expressing some genes at a higher level, leading to a sparse but high overall expression level (Extended Data Fig. 9e). However, using *in situ* hybridization on a different cohort of MS tissue, we could confirm the finding of an increased number of cells expressing *CDH20* in chronic inactive lesions, and a reduced number of *WWOX*-expressing cells in chronic active lesions (Fig. 4c and Extended Data Fig. 9a–c). Thus, in spite of small numbers and variability in pathological lesion subtypes, our results provide proof of principle that MS lesion subtypes may be identifiable by different markers. Increasing patient and lesion numbers may lead to the identification of specific markers of MS lesions, which will be interesting to correlate with clinical outcome, imaging and cell-type-specific effects of MS-risk single nucleotide polymorphisms. These differences may even provide potential future targets for positron-emission tomography biomarkers to identify different MS lesion types *in vivo*.

Our findings clearly illustrate the power of snRNA-seq for the neuropathological analysis of human diseases, and we predict that the widespread use of this technology at scale will greatly enhance our understanding of chronic neurological diseases and lead to revised classifications, improved diagnostic accuracy and new markers. Furthermore, our data show the need to re-evaluate current approaches for discovering regenerative therapies in MS. These are based on the assumption that enhancing the differentiation of resident OPCs to oligodendrocytes that express myelin genes and proteins will lead to enhanced remyelination in progressive MS. Our results show that this is oversimplistic for two reasons. First, the notable pathology we observe is not a failure of differentiation to the myelin-gene-expressing oligodendrocyte, but is instead the loss of the Oligo1 population (which we predict to be fully mature and stable oligodendrocytes) and the skewing of the differentiation program to other subclasses of mature oligodendrocytes with different transcriptional signatures. These new oligodendrocyte subclasses may therefore have important functional differences in their ability to provide metabolic support or, in the case of ImOLG, contribute to the inflammatory pathology. Identification of these functional differences and strategies to restore healthy oligodendrocyte heterogeneity should be a major future focus in MS research. Second, our results showing depletion of not only OPCs but also the intermediate Oligo6 population, and increased expression of myelin genes in mature oligodendrocytes in MS, may suggest that subsets of mature oligodendrocytes contribute to remyelination. This is consistent with retrospective carbon-14-based birth-dating in patients with MS²² and electron microscopy in large animal models²³, but in sharp contrast to rodents in which remyelination is driven entirely by the recruitment and differentiation of resident OPCs. This highlights the difficulties in extrapolation from rodent to human and further emphasizes the power of studying human pathology at a single-cell level to inform appropriate therapeutic approaches.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-0903-2>.

Received: 17 July 2018; Accepted: 15 January 2019;

Published online: 23 January 2019

1. Bodini, B. et al. Dynamic imaging of individual remyelination profiles in multiple sclerosis. *Ann. Neurol.* **79**, 726–738 (2016).
2. Bechler, M. E., Byrne, L. & ffrench-Constant, C. CNS myelin sheath lengths are an intrinsic property of oligodendrocytes. *Curr. Biol.* **25**, 2411–2416 (2015).
3. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
4. Falcão, A. M. et al. Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nat. Med.* **24**, 1837–1844 (2018).
5. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
6. Lassmann, H., Raine, C. S., Antel, J. & Prineas, J. W. Immunopathology of multiple sclerosis: report on an international meeting held at the Institute of Neurology of the University of Vienna. *J. Neuroimmunol.* **86**, 213–217 (1998).
7. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
8. Cui, Q. L. et al. Oligodendrocyte progenitor cell susceptibility to injury in multiple sclerosis. *Am. J. Pathol.* **183**, 516–525 (2013).
9. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
10. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
11. Marques, S. et al. Transcriptional convergence of oligodendrocyte lineage progenitors during development. *Dev. Cell* **46**, 504–517.e7 (2018).
12. Lock, C. et al. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat. Med.* **8**, 500–508 (2002).
13. Baranzini, S. E. et al. Transcriptional analysis of multiple sclerosis brain lesions reveals a complex pattern of cytokine expression. *J. Immunol.* **165**, 6576–6582 (2000).
14. Chabas, D. et al. The influence of the proinflammatory cytokine, osteopontin, on autoimmune demyelinating disease. *Science* **294**, 1731–1735 (2001).
15. Zeis, T., Howell, O. W., Reynolds, R. & Schaeren-Wiemers, N. Molecular pathology of multiple sclerosis lesions reveals a heterogeneous expression pattern of genes involved in oligodendroglogenesis. *Exp. Neurol.* **305**, 76–88 (2018).
16. Dutta, R. & Trapp, B. D. Gene expression profiling in multiple sclerosis brain. *Neurobiol. Dis.* **45**, 108–114 (2012).
17. Boyd, A., Zhang, H. & Williams, A. Insufficient OPC migration into demyelinated lesions is a cause of poor remyelination in MS and mouse models. *Acta Neuropathol.* **125**, 841–859 (2013).
18. Chang, A., Nishiyama, A., Peterson, J., Prineas, J. & Trapp, B. D. NG2-positive oligodendrocyte progenitor cells in adult human brain and multiple sclerosis lesions. *J. Neurosci.* **20**, 6404–6412 (2000).
19. Lucchinetti, C. et al. A quantitative analysis of oligodendrocytes in multiple sclerosis lesions. A study of 113 cases. *Brain* **122**, 2279–2295 (1999).
20. de Groot, M. et al. Changes in normal-appearing white matter precede development of white matter lesions. *Stroke* **44**, 1037–1042 (2013).
21. Huynh, J. L. et al. Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat. Neurosci.* **17**, 121–130 (2014).
22. Yeung, M. S. Y. et al. Dynamics of oligodendrocyte generation in multiple sclerosis. *Nature* <https://doi.org/10.1038/s41586-018-0842-3> (2019).
23. Duncan, I. D. et al. The adult oligodendrocyte can participate in remyelination. *Proc. Natl Acad. Sci. USA* **115**, E11807–E11816 (2018).

Acknowledgements We thank the MS Society UK Tissue Bank and the MRC Sudden Death and MS brain banks for post-mortem brain tissue, Advanced Cell Diagnostics for their help with BaseScope, T. Jimenez-Beristain, A. Nanni and A. Moshref for support, Eukaryotic Single Cell Genomics Facility (ESCGF) at Science for Life Laboratory, L. Wigge (Wallenberg Advanced Bioinformatics Infrastructure (WABI) Long Term Bioinformatic Support at SciLifeLab), the National Genomics Infrastructure, and M. Prasad and F. Koechl at F. Hoffmann-La Roche for providing assistance with snRNA-seq, and B. Vernay, E. O'Duibhir and M. Vermeren (CRM) for imaging support. The bioinformatics computations were performed at Swedish National Infrastructure for Computing (SNIC) at UPPMAX, Uppsala University. Funding: S.J.: European Union, Horizon 2020, Marie-Sklodowska Curie Actions EC no. 789492; C.ff.-C.: Wellcome Trust Investigator award; A.W.: UK Multiple Sclerosis Society, F. Hoffmann-La Roche; E.A.: European Union, Horizon 2020, Marie-Sklodowska Curie Actions, grant SOLO no. 794689; A.M.F.: European Committee for Treatment and Research of Multiple Sclerosis; G.C.-B.: European Union Horizon 2020/European Research Council Consolidator Grant EPISCOPE no. 681893, Swedish Research Council (no. 2015-03558), Swedish Brain Foundation (no. FO2017-0075), Swedish Cancer Society (Cancerfonden, CAN2016/555), Stockholm City Council (grant 20170397), Ming Wai Lau Centre for Reparative Medicine, F. Hoffmann-La Roche.

Reviewer information *Nature* thanks Ed Lein, Klaus-Armin Nave and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.J., E.A., A.M.F., I.K., D.M., C.ff.-C., A.W. and G.C.-B. designed all experiments. S.J. and A.M.F. performed snRNA-seq sample preparation. E.A., D.v.B. and K.W.L. performed the computational analysis of the snRNA-seq data; S.J. performed validation experiments, with assistance of A.W. and I.K.; S.J., E.A., C.ff.-C., A.W. and G.C.-B. wrote the manuscript with input from the co-authors. C.ff.-C., A.W. and G.C.-B. oversaw all aspects of the study.

Competing interests D.M. and I.K. are employees at F. Hoffmann-La Roche.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-019-0903-2>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-0903-2>.

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

Correspondence and requests for materials should be addressed to C.ff.-C., A.W. or G.C.-B.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Human donor tissue. Post-mortem unfixed fresh-frozen tissue and formalin-fixed paraffin-embedded (FFPE) tissue were obtained from the UK Multiple Sclerosis Tissue Bank via a UK prospective donor scheme with full pre-mortem consent and with full ethical approval by MREC/02/2/39 (UK Ethics Committee) and 2016/589-31 (Regionala Etiskprövningssnämnden, Stockholm). For the snRNA-seq, we used white matter regions from fresh-frozen tissue sections for both controls (four men and one woman) and NAWM and MS lesions (three men and one woman). Controls (Supplementary Table 1): five samples from five different donors; NAWM: three samples from three patients; chronic active: four samples from four patients; active: three samples from two patients; chronic inactive: three samples from three patients; remyelinated: two samples from two patients. For the *in situ* validation on FFPE tissue sections, we used a total of 11 control (5 men and 6 women) and 15 MS (7 men and 8 women) tissue samples from different donors. The ages of the control and MS donors did not differ significantly (control frozen: 58.0 ± 17.5 years; MS frozen: 46.8 ± 8.4 years; control FFPE: 57.7 ± 12.3 years, MS FFPE: 53.5 ± 9.1 years; one-way ANOVA, $P = 0.3724$, $F_{(3,31)} = 1.079$; Tukey's multiple comparison test: control frozen versus MS frozen: $P = 0.4769$, control frozen versus control FFPE: $P > 0.9999$, control frozen versus MS FFPE: $P = 0.8713$, MS frozen versus control FFPE: $P = 0.3775$, MS frozen versus MS FFPE: $P = 0.7307$, control FFPE versus MS FFPE: $P = 0.7889$). Data are mean \pm s.d.

Isolation of nuclei. Nuclei were isolated from fresh-frozen 10- μm sections as previously described²⁴ with modifications. The regions of interest were macro-dissected with a scalpel blade, lysed in nuclei isolation buffer (10 mM Tris-HCl pH 8.0, 0.25 M sucrose, 5 mM MgCl₂, 25 mM KCl, 0.1% Triton-X) with 0.1 mM dithiothreitol (DTT) and 0.4 U μl^{-1} RNase inhibitors freshly added before use and homogenized with a Dounce homogenizer. The suspension was filtered through a 30- μm strainer and centrifuged for 10 min at 1,000g. The pellet was re-suspended in 400 μl cold PBS with 0.4 U μl^{-1} RNase inhibitors and 310 μl of solution mixed with 90 μl of debris removal solution (Miltenyi Biotech), overlaid with 400 μl of cold PBS with 0.4 U μl^{-1} RNase inhibitors and centrifuged for 10 min at 3,000g. The supernatant was removed, the pellet washed with cold PBS with 0.4 U μl^{-1} RNase inhibitors and re-suspended in PBS 0.5% BSA with 0.4 U μl^{-1} RNase inhibitors. The remaining 90 μl was diluted with 180 μl of cold PBS, 0.75% BSA, 0.4 U μl^{-1} RNase inhibitors, filtered through a decreasing cell strainer size (30–10 μm) and centrifuged for 5 min at 1,000g. The pellet was re-suspended in 25 μl PBS 0.5% BSA with 0.4 U μl^{-1} RNase inhibitors. The two pellets were combined 1:4 (filtered:debris removed, respectively) for further 10x loading.

Single-nuclei preparation for 10x loading. Wash buffer (25 μl ; 10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl₂, 0.005% NP40) with 0.2 U μl^{-1} RNase inhibitors was added to each nuclei suspension, gently mixed and incubated for 5 min on ice. The suspension was centrifuged for 5 min at 1,000g and the pellet gently re-suspended in PBS, 2% BSA with 0.2 U μl^{-1} RNase inhibitors. For quantification, the nuclei were stained with Hoechst (5 $\mu\text{g ml}^{-1}$) and counted in a haemocytometer. A total of 8,000 estimated nuclei for each sample was loaded on the 10x Single Cell A Chip, although a much lower number of nuclei was recovered after sequencing (Supplementary Table 2).

cDNA library preparation. cDNA libraries have been prepared using the Chromium Single Cell 3' Library and Gel Bead kit v2 (120267) according to the manufacturer's instructions.

Pre-processing and clustering of MS and control samples. The 20 samples were aligned with Cellranger v.2.1.1 with reference genome GRCh38-1.2.0. Then, each of the output filtered UMI count matrixes was used as input for Velocyto²⁵ with the parameters, 'velocyto' 'run10x -m repeatMasker_filtered_UMI_count_matrixes' 'GRCh38-1.2.0_genes.gtf'. The repeatmasker track was download from UCSC tables. Velocyto only considers uniquely mapped reads from cellranger output UMI matrixes and reads that align to both exonic regions and intronic regions. The new UMI count matrixes were exported from loom file format to R object format with the Velocyto²⁵ R package. For each of the samples, we combined the spliced and unspliced count matrixes to get a matrix of 33,692 genes across 35,753 cells. The final aggregate UMI matrices were used for all the downstream analyses. We checked quality metrics and removed cells with less than 200 genes and a total count below 500, and we excluded genes without a count above 1 in at least 3 cells.

The following data processing was carried out with Seurat⁷ (v.2.1). For each of the 20 samples, we initially set up a first filter of min.cells = 3 and min.genes = 200 per sample, filtered by number of UMIs (>6,000), genes (<200) and mitochondrial percentage (>0.20). The distribution for gene, UMIs and reads mapped to mitochondrial genome were visually inspected and used for quality assurance. Post-processed matrices were then log-normalized individually with a scale of factor of 10,000, followed by regressing intercellular variation in gene expression by UMI counts and batch number, and scaling of the gene expression. Highly variable genes were set up as the union from the top 1,000 highly variable genes from each sample, resulting in 4,361 genes.

After quality filtering, 17,799 nuclei and 21,581 genes remained. The shared-nearest neighbour graph was constructed on a cell-to-cell distance matrix from top 15 aligned canonical correlation vectors. The shared-nearest neighbour graph with different resolution was used as an input for the smart local moving algorithm to obtain cell clusters, and visualized with t-SNE. We performed the analysis in three different resolutions: 0.8, 2.0 and 4.0. On the basis of differentially expressed genes, identified by Wilcoxon rank sum test, with parameters min.pct = 0.25, thresh.use = 0.25, test.use = "wilcox", we manually assigned and verified the consistency of the three different resolutions (Extended Data Fig. 2). On the basis of previous knowledge and consistency within the different resolutions, we selected the final number of clusters between resolutions 2.0 and 4.0, which included all the major cell types in the brain and novel cell types⁴, resulting in 23 different clusters.

Oligodendrocyte cell-type assignment. The CCA applied for the control and MS combined analysis minimized inter-sample variability to reduce possible batch effects due to the individual variability or technical performance, leading to intermingled comparable clusters that contained MS and control nuclei. First, oligodendrocyte lineage cell types were identified on the basis of canonical and novel markers³ from differential expression analysis. To verify cell identity, the expression pattern distributions were used in the different Oligo subclusters, and then verified as separate on the basis of the markers. In addition, evidence from mouse single-cell data⁴ has shown oligodendrocytes with an immunological phenotype in a mouse model of experimental autoimmune encephalomyelitis (EAE). The combination of different resolutions allowed us to identify oligodendrocytes expressing immune genes in the human dataset (ImOLG).

To verify the subcluster identity and possible over- or under-clustering, a classification hierarchy was built. This approach places transcriptionally similar clusters close to each other on a tree allowing us to finally define the six Oligo clusters, the OPCs, the COPs and the ImOLG as different separate clusters that were later validated with specific markers.

Dimensional reduction with principal component analysis (PCA), followed by regressing out each of the sample variables, showed segregation of clusters based on patient identity (Extended Data Fig. 7), suggestive of batch effects and individual variability. Thus, we performed CCA, considering all the individual samples as a variable and using the union of the top variable genes from each of the samples to get common—but also specific—variable genes from all samples and discarding cells with higher than 0.5 PCA/CCA variance. CCA allows the alignment of all samples to a common low-dimensional subspace followed by clustering⁷, showing that nuclei cluster more according to cell type than sample identity (Fig. 1a and Extended Data Fig. 3).

Comparison of human and mouse oligodendroglia. The normalized expression matrix that was previously described⁴ was retrieved and mouse mm10 Genesymbol IDs were extracted and combined with GRCh38 ENSEMBL geneIDs from Biomart²⁶. We recovered a final matrix with unique GRCh38 gene symbols renamed from mm10. For the comparison analysis we combined in a single matrix the six Oligo clusters, OPCs, COPs and ImOLG from the human dataset with the renamed clusters from the EAE mouse dataset. Both datasets included only oligodendrocyte-lineage cell clusters that combined MS, or EAE in mouse, and controls. The combined 900 most variable genes as previously described²⁷ from all the mouse and human oligodendrocytes were used to classify the cell types. The dataset similarity analysis was performed with an unsupervised classifying approach to find the most similar cell types, using as training and testing both datasets, and with a top hits threshold of ≥ 0.7 mean area under the receiver operator characteristic curve (AUROC)²⁷ score.

Integration of different human snRNA-seq datasets. UMI count matrices from the human cerebellar hemisphere (24,580 genes, 5,204 nuclei), frontal cortex (24,654 genes, 10,319 nuclei) and visual cortex (32,693 genes, 1,386 nuclei) were retrieved from the GEO (accession GSE97930), and pre-processed independently. First, nuclei that had no annotation were discarded (1 cell from frontal cortex UMI count matrix). We performed a quality control procedure as described above. Nuclei with a number of genes $> 3,100$ (cerebellar hemisphere, frontal and visual cortex) and < 350 (frontal and visual cortex), and number of UMIs $> 5,600$ (cerebellar hemisphere, frontal and visual cortex) were considered as low quality or outliers, and were thereafter removed from downstream analysis. After quality controls, 5,199, 9,557 and 18,645 nuclei remained (cerebellar hemisphere, frontal and visual cortex). Post-processed matrices were then log-normalized individually with a scale of factor of 10,000, followed by regressing inter-cellular variation in gene expression by UMI counts and batch number, and scaling of the gene expression. Then, CCA⁷ was performed for each region. The first 17 canonical correlation components were chosen. There were 465 nuclei in the 17 correlation components with PCA or CCA variance more than 0.5, and these were thus considered regional or batch specific and removed for dataset alignment. After aligning regional expression in the first 17 correlation components, clustering, visualization and oligodendrocyte sub-setting were performed.

UMI count matrices of human archived brain samples were retrieved from https://portals.broadinstitute.org/single_cell/study/dronec-seq-single-nucleus-rna-seq-on-human-archived-brain#study-summary. Hippocampus (HIP) (10,326 genes, 5,433 nuclei) and prefrontal cortex (PFC) (10,326 genes, 9,530 nuclei) expression matrices were separated and underwent quality control. Cells with a number of genes $>4,000$ and <400 (PFC) and <350 (HIP), number of UMIs $>7,500$ (PFC and HIP), and mitochondrial percentage <0.05 (PFC) and >0.15 (HIP) were considered as low quality or outliers and were thereafter removed from downstream analysis. A total of 6,062 PFC nuclei and 4,765 HIP nuclei remained. Each regional dataset was then log-normalized with a scale of factor of 10,000, followed by regressing intercellular variation in gene expression by UMI counts and mitochondrial percentage, and scaling of the gene expression. Highly variable genes were identified as with the Lake et al.⁹ dataset, but with the high end cut-off of 4 for average expression, and a union of 1,521 variable genes was used for canonical correlation analysis. CCA or PCA variance, and dataset low-dimensional subspace alignments, followed by clustering and oligodendrocytes linkage sub-setting, were performed as above with the first 11 correlation components with 9,702 nuclei in total after discarding cells with higher than 0.5 PCA or CCA variance.

Dataset integration. Oligodendroglial subsets from our controls, and the Lake et al.⁹ and Habib et al.¹⁰ datasets were combined by performing a CCA with the union of top 1,000 highly variable genes from each dataset, and then 11 correlation components were aligned after discarding data-specific nuclei. Differentially expressed genes that were conserved among the datasets were identified by first performing individual within-dataset Wilcoxon rank sum tests, followed by ranking genes according to a unified combined Fisher's *P* value. The resulting clusters were found under resolution 1.0.

Clustering based on PCA. Seurat was also used for PCA of the 20 MS and control samples. The filtered expression matrix, as previously described, was log-normalized with a scale factor of 10,000, scaled and regressed on the number of UMIs and sample identity. PCA was run on highly variable genes, which were identified as previously described. Fifteen principal components were used and the HNN graph was constructed based on the Euclidean distance in PCA space, in which the clusters were then identified using Louvain algorithm. The clusters were visualized using t-SNE. Clustering was run in three different resolutions, 0.8, 2.0 and 4.0, to be comparable with the clusters obtained with the CCA.

Spatial gene-filtering and pseudo-ordering. Cells were ordered, and lineages were approximated using a previously published pipeline¹¹. In brief, single nuclei were filtered so that each nucleus contained at least 500 UMI counts, and at least 400 genes. We then used spatially correlating gene selection on the diffusion mapping²⁸ obtained transition matrix. Subsequently, we reduced the high-dimensional space using non-negative matrix factorization²⁹, of which the ideal ranks are estimated using a measure of mutual information across the obtained components. The chosen rank was obtained by selecting the rank number for which the calculated joined mutual information no longer highly decreases upon increasing rank. The non-negative matrix is then transformed in a transition space using diffusion mapping for which then lineages are calculated. See <https://github.com/Castelo-Branco-lab/GeneFocus> for recent code.

Differential expression analysis between MS lesions and controls. For differential expression analysis, MAST was used³⁰. All oligodendrocyte lineage cells were included in the analysis. An FDR value of 0.05 was taken, and genes with a fold change (expressed logarithmically) of at least 5 were selected to be included in the plot in Fig. 4a. Proportional expression was calculated by taking the mean expression value of the nuclei, leading to a heavily 0 biased threshold. Nuclei expressing higher than this threshold are considered to be expressing nuclei, the proportion of expressing nuclei was then calculated on a per gene basis.

GO analysis. For GO analyses, the most significantly differentially expressed genes from the snRNA-seq experiment of each oligodendrocyte subcluster were selected (adjusted *P* ≤ 0.05 , fold change (expressed logarithmically) ≥ 0.5). GO and pathway analysis was performed with the ClueGO (v.2.5.2) plug-in Cytoscape (v.3.7.0)³¹ with settings, GO Biological process (4 September 2018) and REACTOME pathways (4 September 2018), showing only pathways with *P* ≤ 0.05 . Default settings and GO fusion were used; for subclusters with more than 200 significantly regulated genes (OPCs and COPs), a minimum of 7 genes per cluster was used; for subclusters with less than 50 significantly regulated genes (Oligo4 and Oligo2), a minimum of 2 genes per cluster was used; and for all other subclusters, a minimum of 5 genes per cluster was used.

Immunohistochemistry. FFPE sections (4 μm) were deparaffinized in decreasing concentrations of ethanol, and antigen retrieval was performed in antigen unmasking solution (Vector Laboratories, H-3300) for 10 min. For colorimetric labelling, sections were washed in PBS, blocked for 30 min at room temperature with PBS, 0.5% Triton-X (PBS-T) and 10% heat-inactivated horse serum (HIHS, blocking buffer). Primary antibody incubation was performed overnight at 4 °C in blocking buffer. Sections were washed in PBS and incubated with a horseradish peroxidase (HRP)- or alkaline phosphatase (AP)-labelled secondary antibody appropriate

for the respective species for 2 h at room temperature. Colour reaction was performed using DAB or VectorBlue reaction kits (Vector Laboratories, SK-4100 and SK-5300, respectively). Sections were washed and mounted. For fluorescent labelling, deparaffinized sections were incubated with autofluorescence eliminator reagent (Millipore, 2160) for 1 min and washed with TBS 0.001% Triton-X (wash buffer). Endogenous peroxidases were quenched with 3% H₂O₂ for 15 min at room temperature, washed and blocked for 30 min at room temperature with TBS 0.5% Triton-X (TBS-T), 10% HIHS (blocking buffer 2). Primary antibody incubation was performed overnight at 4 °C in blocking buffer 2. Fluorophore reaction was performed using thymidine reaction kits for fluorescein, cyanine 3 and cyanine 5 (Perkin Elmer, NEL741B001KT, NEL744B001KT and NEL745B001KT, respectively). Sections were counterstained using Hoechst (Thermo Fisher, 62249; 1:1,000), washed and mounted. The following primary antibodies were used: rabbit anti-OLIG2 (Atlas, HPA003254; 1:100), goat anti-OLIG2 (R&D Systems, AF2418; 1:100), rabbit anti-OLIG1 (Abcam, ab68105; 1:100), rabbit anti-MRF (Millipore, ABN45, 1:100), rabbit anti-OPALIN (Abcam, ab121425, 1:100), rabbit anti-SOX6 (Millipore, AB5805, 1:100) and goat anti-KLK6 (Life Technologies, PA547239, 1:100), mouse anti-CNP (Atlas, AMAb91072, 1:1,000), rat anti-NOGO (R&D Systems, MAB3098, 1:100). The specificity of our MRF antibody was validated by western blot as well as a combination of mRNA and protein labelling analysis (Extended Data Fig. 8). The following secondary antibodies were used: Vector Laboratories, rabbit-HRP IgG (MP-7401), rabbit-AP IgG (MP-5401), goat-HRP (MP-7405), mouse-HRP IgG (MP-7402), mouse-AP IgG (MP-5402).

BaseScope mRNA detection. BaseScope mRNA detection was performed according to the manual, using the RNAScope pretreatment and wash buffer reagents (ACD, 322380 and 310091, respectively) and BaseScope Red and BaseScope duplex detection kits (ACD, 322910 and 3223810, respectively). FFPE sections (4 μm) were deparaffinized twice for 5 min in xylene and twice for 2 min in ethanol. Sections were dried and incubated with H₂O₂ for 10 min. Pretreatment was performed for 45 min with pretreatment buffer, followed by incubation for 3 min in ethanol, and slides were dried overnight. Protease treatment was performed using protease IV for 1 h at 40 °C, and the protease was refreshed after 15 min. Sections were washed in deionized water and probes were incubated for 2 h at 40 °C. Sections were washed in wash buffer and the colour reaction was performed according to the user manual with the following adjustments: for the single BaseScope detection, the AMP5 step was increased to 45 min. For the duplex assay, the steps Amp7 and Amp11 were increased to 1 h. All the probes have been designed and produced by Advanced Cell Diagnostics. Probes against the following human genes were used: channel 1 probes: BCAN, CDH20, WWOX, KLK6, CLDND1, MRF, CD74 and OPALIN. Channel 2 probes: LURAP1L-AS1 and CDH20. For the single assay, subsequent IHC was performed as described above. For the duplex assay, sections were counterstained for 1 min with haematoxylin (Scientific Laboratories, GHS132-IL), and the blue reaction was performed using 0.02% ammonia water. Sections were dried at 60 °C and mounted.

Image analysis. IHC was performed automated with image analysis software to exclude any bias and to be blinded. In the other cases, quantification has been validated by two independent people. Brightfield images have been acquired using a Widefield Observer (Zeiss) inverted microscope and a Vectra Polaris (Perkin Elmer) slide scanner. Fluorescent images (*z*-stacks) have been acquired using a confocal microscope (Leica TCS SP8). Image analysis has been performed using open source Fiji³² and QuPath³³ imaging software. For cell quantification of fluorescent images and brightfield images in which no automated quantification was possible, a minimum of eight randomly chosen regions of equal dimensions per patient and region were acquired. Total cell numbers per mm² were calculated based on the image dimensions. For fluorescent images, *z*-stacks were collapsed to a maximum intensity projection and the number of cells was quantified using the Fiji cell counter plugin. For quantification of the single-channel brightfield images and the BaseScope IHC double-positive images, the number of cells was also quantified using the Fiji cell counter plugin. The average of these different regions was taken and considered as *n* = 1. Where possible, brightfield images were quantified automatically. Therefore, whole slides were scanned using a slide scanner. Using QuPath, a minimum of four random regions per sample and condition were annotated, and cells within these regions were quantified using the automated ‘positive cell detection’ plugin.

For the quantification of the duplex BaseScope, random brightfield images were acquired in the white matter of control donors. All cells with a positive BaseScope signal were quantified and represent 100% of labelled cells; the percentages of single- or double-positive cells were determined.

For the quantification of the differences in mRNA expression between MS lesions, whole sections were scanned with a slide scanner. Demyelinated lesions were detected by the absence of IHC staining for CNP. Within these regions, between 4 and 11 random regions of equal size were annotated using QuPath and the number of BaseScope-positive dots was quantified. The average of the random regions within 1 lesioned area was considered as *n* = 1.

Western blot. Fresh frozen human brain sections ($10\text{ }\mu\text{m}$) were lysed in RIPA buffer (Thermo Scientific, 89900), sonicated, centrifuged for 10 min at 13,000g and the supernatant was then collected. Ten microlitres of supernatant per lane was diluted in RIPA and Laemmli-buffer and incubated for 5 min at 96°C . Proteins were separated on an SDS-PAGE gel (Bio-Rad, 161-1176) together with a protein ladder (Bio-Rad kaleidoscope standard, 161-0375). Proteins were blotted onto a PVDF membrane in a wet blotting chamber with transfer buffer ($1\times$ Tris-glycine buffer, 20% methanol; 2 h at 400 mA). Membranes were washed with $1\times$ TBS, 0.01% Tween (TBS-Tween) and blocked for 1 h with TBS-Tween, 5% milk. The primary antibody (rabbit anti-MRF, Millipore, ABN45; 1:500) was incubated in blocking buffer overnight at 4°C . The membrane was washed and secondary antibody (anti-rabbit HRP, Vector Laboratories, MP-7401; 1:10,000) was incubated for 1 h at room temperature in blocking buffer. The membrane was washed and incubated for 5 min with ECL solution (Thermo Scientific, 1863031). Proteins were visualized using X-ray film.

Statistics. Statistical analysis was performed using GraphPad Prism 7. In Fig. 1g, n represents the number of biologically independent individuals. We used $n = 4$ for *LURAP1L-AS1⁺CDH20⁺*, and $n = 3$ for the other combinations of oligodendrocyte subclass markers. No statistics were applied in Fig. 1. In Fig. 3a, n represents the number of different (that is, biologically independent) donors. We used $n = 5$ control and $n = 9$ patients with MS. Haematoxylin: one-way ANOVA, $F_{(2,20)} = 34.8$, $P < 0.0001$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.7445$, control WM versus lesions: $P < 0.0001$, NAWM versus lesions: $P < 0.0001$. OLIG2 per mm²: one-way ANOVA, $F_{(2,20)} = 11.3$, $P = 0.0005$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.6298$, control WM versus lesions: $P = 0.0013$, NAWM versus lesions: $P = 0.0029$. OLIG2 percentage: one-way ANOVA, $F_{(2,20)} = 3.553$, $P = 0.0478$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.9687$, control WM versus lesions: $P = 0.1707$, NAWM versus lesions: $P = 0.0523$. In Fig. 3b (SOX6), n represents the number of different donors. We used $n = 4$ control and $n = 5$ NAWM and lesions. One-way ANOVA, $F_{(2,11)} = 50.42$, $P < 0.0001$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.0032$, control WM versus lesions: $P < 0.0001$, NAWM versus lesions: $P = 0.0003$. For Fig. 3c (OPALIN), we used $n = 3$ for control and $n = 5$ for NAWM and lesions. One-way ANOVA, $F_{(2,10)} = 147.9$, $P < 0.0001$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.0002$, control WM versus lesions: $P < 0.0001$, NAWM versus lesions: $P < 0.0001$. In Fig. 4c (CDH20 BaseScope), n represents the number of different lesions from separate individuals. We used $n = 5$ for active lesion, $n = 7$ for chronic inactive lesions and $n = 3$ for chronic active lesions in a total of 7 patients with MS. One way ANOVA, $F_{(2,12)} = 5.473$, $P = 0.0205$. Tukey's multiple comparison test: active versus chronic inactive lesions: $P = 0.0214$, active versus chronic active lesions: $P = 0.8439$, chronic inactive versus chronic active lesions: $P = 0.1368$. For Extended Data Fig. 4b–e, we used $n = 3$ different donors to validate the co-labelling of each marker. For Extended Data Fig. 4g (OPALIN bins), we used $n = 3$ different donors for each group. One-way ANOVA, $F_{(2,6)} = 73.89$, $P < 0.0001$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.0007$, control WM versus lesions: $P < 0.0001$, NAWM versus lesions: $P = 0.0093$. In Extended Data Fig. 8a (BCAN), n represents the number of different donors. We used $n = 4$ control, $n = 6$ NAWM and $n = 5$ lesions. One-way ANOVA, $F_{(2,12)} = 38.39$, $P < 0.0001$, Tukey's multiple comparison test: control WM versus NAWM: $P < 0.0001$, control WM versus lesions: $P < 0.0001$, NAWM versus lesions: $P = 0.0634$. In Extended Data Fig. 8b (KLK6), n represents the number of different donors. We used $n = 4$ control and $n = 5$ NAWM and lesions. One-way ANOVA, $F_{(2,11)} = 6.742$, $P = 0.0123$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.2150$, control WM versus lesions: $P = 0.2621$, NAWM versus lesions: $P = 0.0095$. In Extended Data Fig. 8d (MRF), n represents the number of different donors. We used $n = 6$ control and NAWM and $n = 7$ lesions. One-way ANOVA, $F_{(2,16)} = 44.63$, $P < 0.0001$. Tukey's multiple comparison test: control WM versus NAWM: $P = 0.0015$, control WM versus lesions: $P < 0.0001$, NAWM versus lesions: $P = 0.0004$. For Extended Data Fig. 9a, the individual number of quantified mRNA molecules per field per patient ($n = 7$) are shown. We used the following number of fields: MS235: $n = 10$

for active and chronic inactive lesions, MS200: $n = 4$ for active, CI and CA lesions, MS249: $n = 4$ for active and $n = 8$ for chronic inactive lesions, MS361: $n = 7$ for active and $n = 10$ for chronic inactive lesions, MS106: $n = 11$ for chronic active and inactive lesions, MS161: $n = 6$ for chronic active and $n = 10$ for chronic inactive lesions, MS300: $n = 7$ for active and $n = 10$ for chronic inactive lesions. No statistics applied. For Extended Fig. 9b (*WWOX*) n represents the number of different lesions from separate individuals. We used $n = 2$ active lesions, $n = 5$ chronic inactive lesions and $n = 4$ chronic active lesions in a total of 5 patients with MS. No statistics applied. For Extended Data Fig. 9c, the individual number of quantified mRNA molecules per field per patient ($n = 5$) is shown. We used the following number of fields: MS245: $n = 8$ for active, $n = 10$ for chronic inactive and $n = 9$ for chronic active lesions, MS361: $n = 6$ for active and $n = 10$ for chronic inactive lesions, MS101: $n = 6$ for chronic inactive and $n = 11$ for chronic active lesions, MS161: $n = 10$ for chronic inactive and $n = 7$ for chronic active lesions, MS296: $n = 11$ for chronic active and $n = 6$ for chronic inactive lesions. No statistics applied.

No statistical methods were used to predetermine sample size. For both the snRNA-seq and the *in situ* validations, we used donor tissue from both sexes randomly distributed in each group (see Supplementary Table 1). The age of the donors was also randomly chosen and distributed between the groups and does not display significant differences (see Supplementary Table 1). For IHC, the fields for image quantification have been chosen randomly throughout the tissue section so as not to introduce any regional bias. The tissue for the snRNA-seq experiment could not be collected blinded, as specific areas had to be chosen. However, the experiment and subsequent data analysis were performed fully automated and hence do not display any bias. Where possible, analysis of IHC was performed fully automated with image analysis software to exclude any bias. In the other cases, quantification was validated by two independent people.

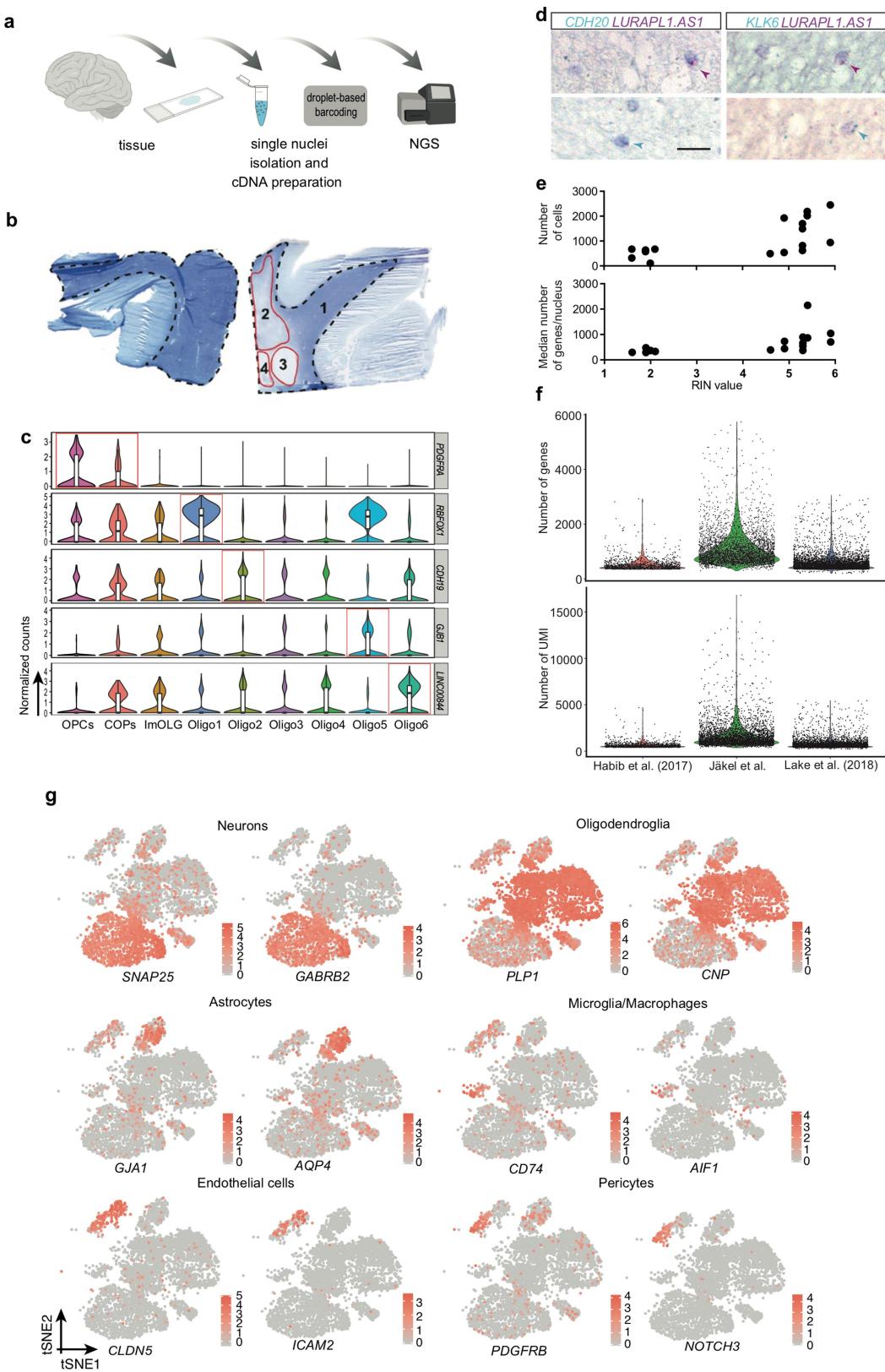
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. All source code and notebooks can be found at the github webpage https://github.com/Castelo-Branco-lab/Jaekel_Agitre_et_al_2018.

Data availability

Sequence data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003412. UMI expression and cell-type annotation tables have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE118257. A browsable webresource is available at <https://ki.se/en/mgb/oligointernode>.

24. Hochgerner, H. et al. STRT-seq-2i: dual-index 5' single cell and nucleus RNA-seq on an addressable microwell array. *Sci. Rep.* **7**, 16327 (2017).
25. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
26. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protocols* **4**, 1184–1191 (2009).
27. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).
28. Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
29. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
30. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
31. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
32. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
33. Bankhead, P. et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

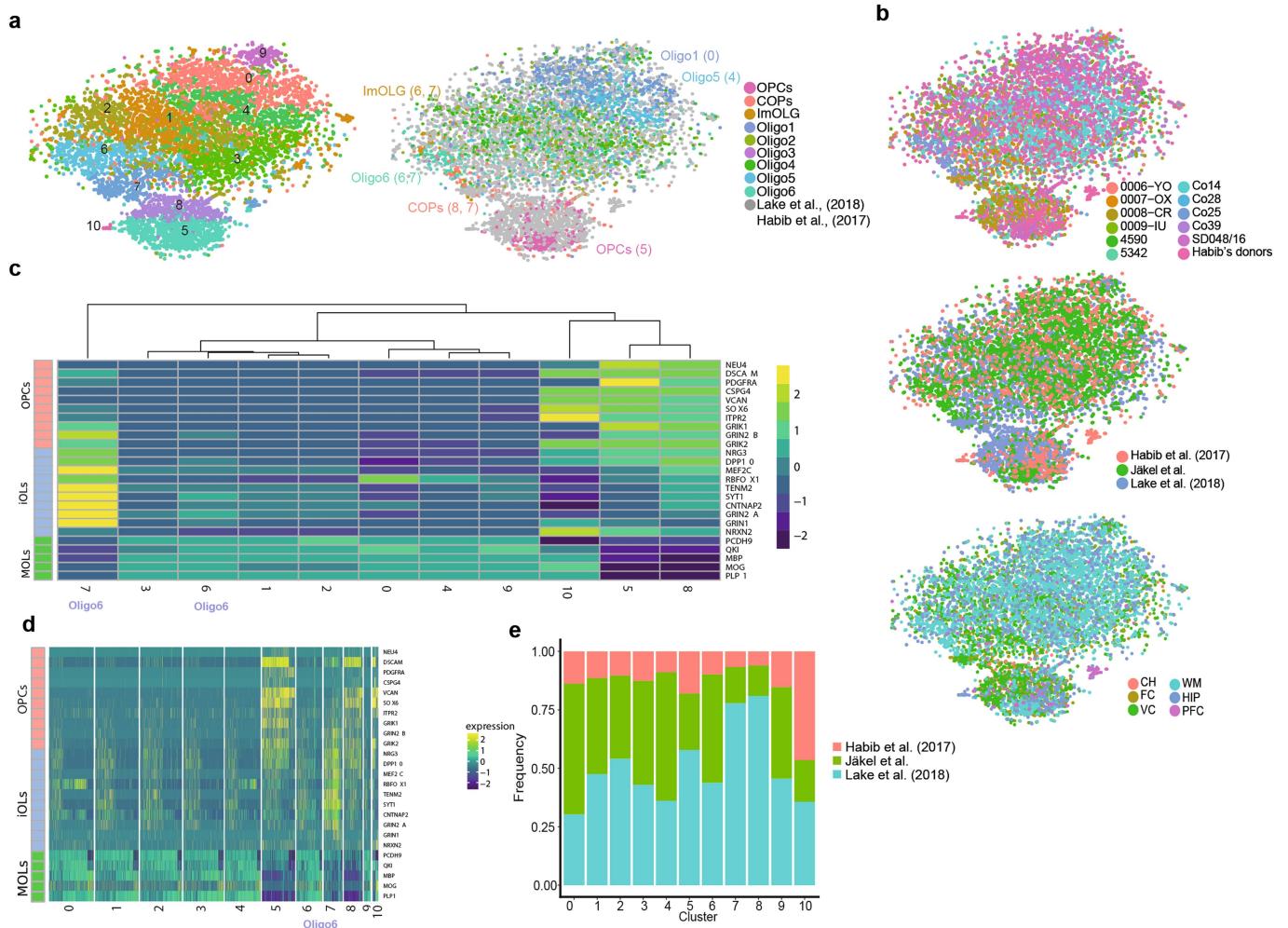


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | snRNA-seq of human post-mortem brain tissue.

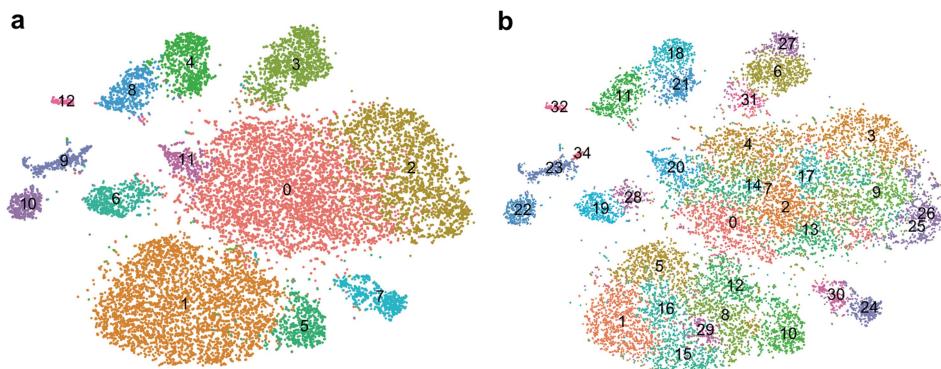
a, Schematic overview of the methodology and workflow used to isolate single nuclei from human white matter, and RNA-seq analysis using Chromium 10x Genomics and Illumina next-generation sequencing (scheme was created with BioRender). **b**, Luxol Fast Blue (LFB) staining of human control (left) and MS (right) brain sections; white matter is outlined with a dotted line. MS brains were divided into NAWM (1) and different lesion types (2–4). **c**, Violin plots of additional markers enriched in specific oligodendrocyte subpopulations showing normalized gene expression (OPC $n = 352$, COP $n = 242$, ImOLG $n = 207$, Oligo1 $n = 1,129$, Oligo2 $n = 1,839$, Oligo3 $n = 775$, Oligo4 $n = 1,579$, Oligo5 $n = 1,167$, Oligo6 $n = 1,484$). Violin plots are centred on the

median with interquartile ranges, and the shape represents cell distribution. **d**, Double *in situ* hybridization (BaseScope) of human control white matter counterstained with haematoxylin. **e**, Correlation between RNA integrity number (RIN) values and the number of genes per nucleus or number of nuclei recovered in individual samples. **f**, Quality control parameters of different human brain oligodendrocyte snRNA-seq datasets showing the individual number of genes (top) and the number of UMIs (bottom) per nucleus ($n = 1,161$ nuclei from Habib et al.¹⁰, $n = 3,998$ control nuclei from this dataset, and $n = 4,873$ nuclei from Lake et al.⁹). **g**, *t*-SNE projections of known cellular markers for the identification of all brain cell clusters in control samples ($n = 6,591$ nuclei).

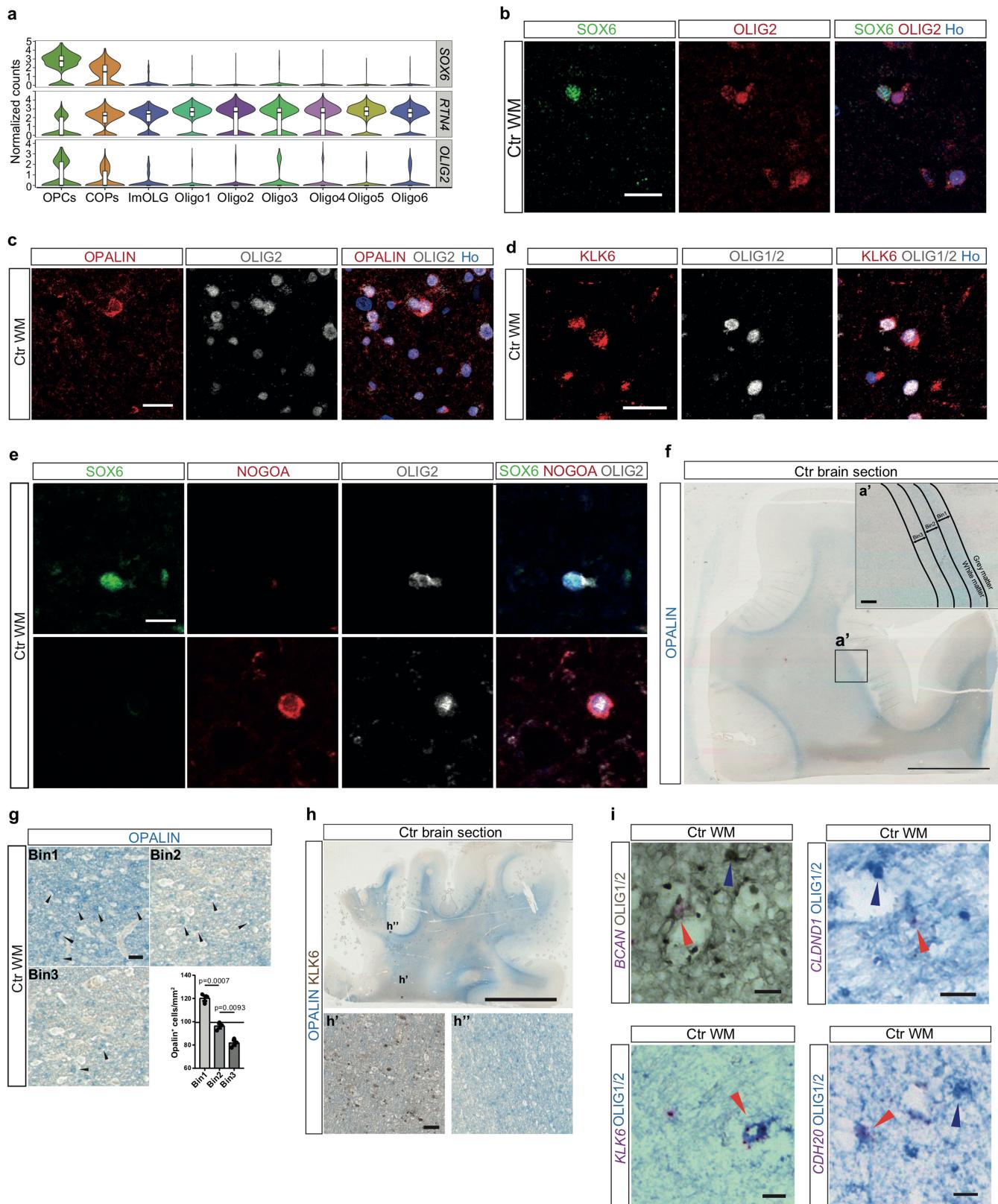


Extended Data Fig. 2 | Combination with other human brain snRNA-seq datasets identifies the Oligo6 subcluster as an intermediate oligodendrocyte state. **a**, t-SNE projections representing oligodendrocyte lineage clusters when performing clustering analysis with the combination of the three datasets (left) and assigning cell identity according to the clusters identified in Fig. 1 (right, in parentheses, the numerical cluster identity with the dataset combination, as indicated in the left t-SNE analysis) (n = number of nuclei; nCluster0 = 1,445; nCluster1 = 1,406; nCluster2 = 1,355; nCluster3 = 1,299; nCluster4 = 1,150; nCluster5 = 1,068; nCluster6 = 828; nCluster7 = 605; nCluster8 = 59;

nCluster9 = 250; and nCluster10 = 28). **b**, t-SNE projections indicating the cell origin when combining the current snRNA-seq dataset with datasets from Habib et al.¹⁰ and Lake et al.⁹ snRNA-seq datasets sorted by different individuals (top), different datasets (middle) and different regions (bottom) ($n = 9,493$ nuclei). **c**, **d**, Heat maps representing expression of genes associated with intermediate states across the oligodendroglial lineage (as defined by Lake et al.⁹) at a cluster (c) and individual nuclei (d) level. **e**, Frequency distribution of identified oligodendroglia between different datasets.



Extended Data Fig. 3 | Seurat CCA clustering of snRNA-seq dataset at different clustering resolutions. a, b, Seurat clustering at a lower (a) and higher (b) resolution than the clustering resolution in Fig. 1 ($n = 17,799$ nuclei derived from 5 control individuals and 4 patients with MS).

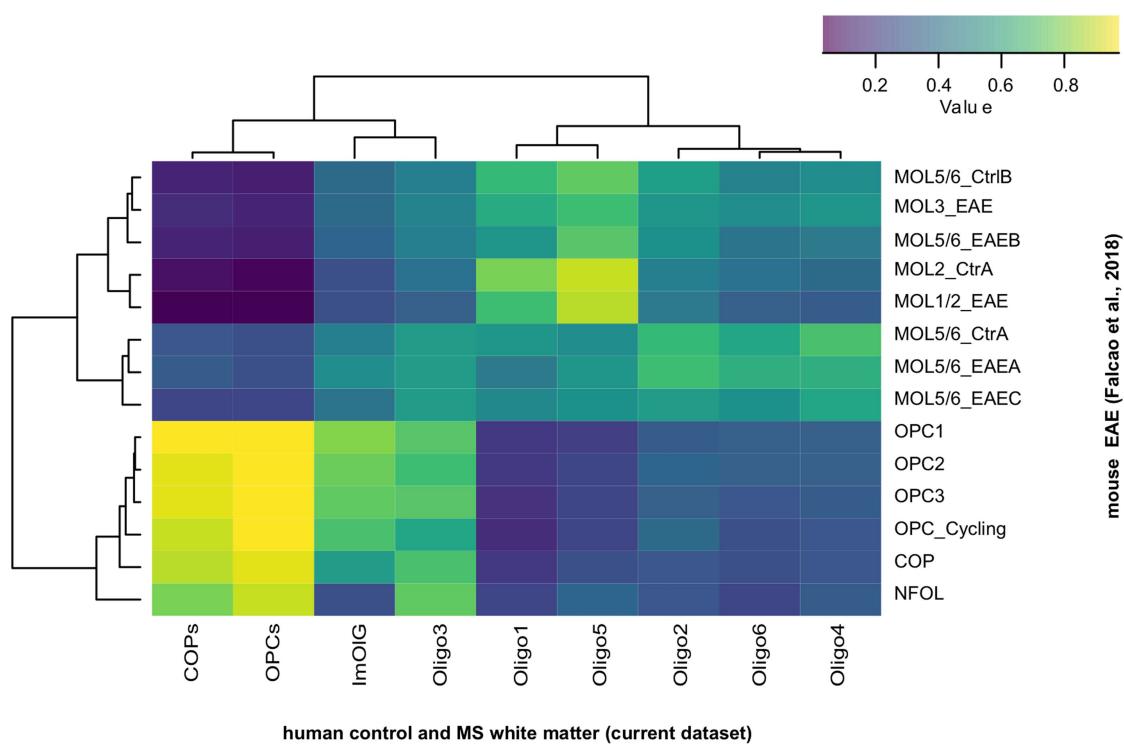


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Validation of oligodendrocyte subcluster markers and regional oligodendrocyte subpopulation distribution in human control brain. **a**, Violin plots showing normalized expression counts of SOX6, *RTN4* (encoding NOGO) and *OLIG2* counts in different oligodendrocyte subpopulations (n = number of nuclei; control OPC $n = 273$; COP $n = 153$; ImOLG $n = 81$; Oligo1 $n = 952$; Oligo2 $n = 388$; Oligo3 $n = 82$; Oligo4 $n = 724$; Oligo5 $n = 393$; Oligo6 $n = 991$). Violin plots are centred on the median with interquartile ranges, and the shapes represent cell distribution. **b**, Colocalization of SOX6 and OLIG2 as a marker for OPCs. Scale bar, 20 μm . **c**, Colocalization of OPALIN and OLIG2 as a marker for Oligo6. Scale bar, 20 μm . **d**, Colocalization of KLK6 and OLIG1/OLIG2 as a marker for Oligo5. **e**, Colocalization of SOX6, NOGO and OLIG2. SOX6 $^+$ OLIG2 $^+$ NOGO $^-$ cells (top) are OPCs,

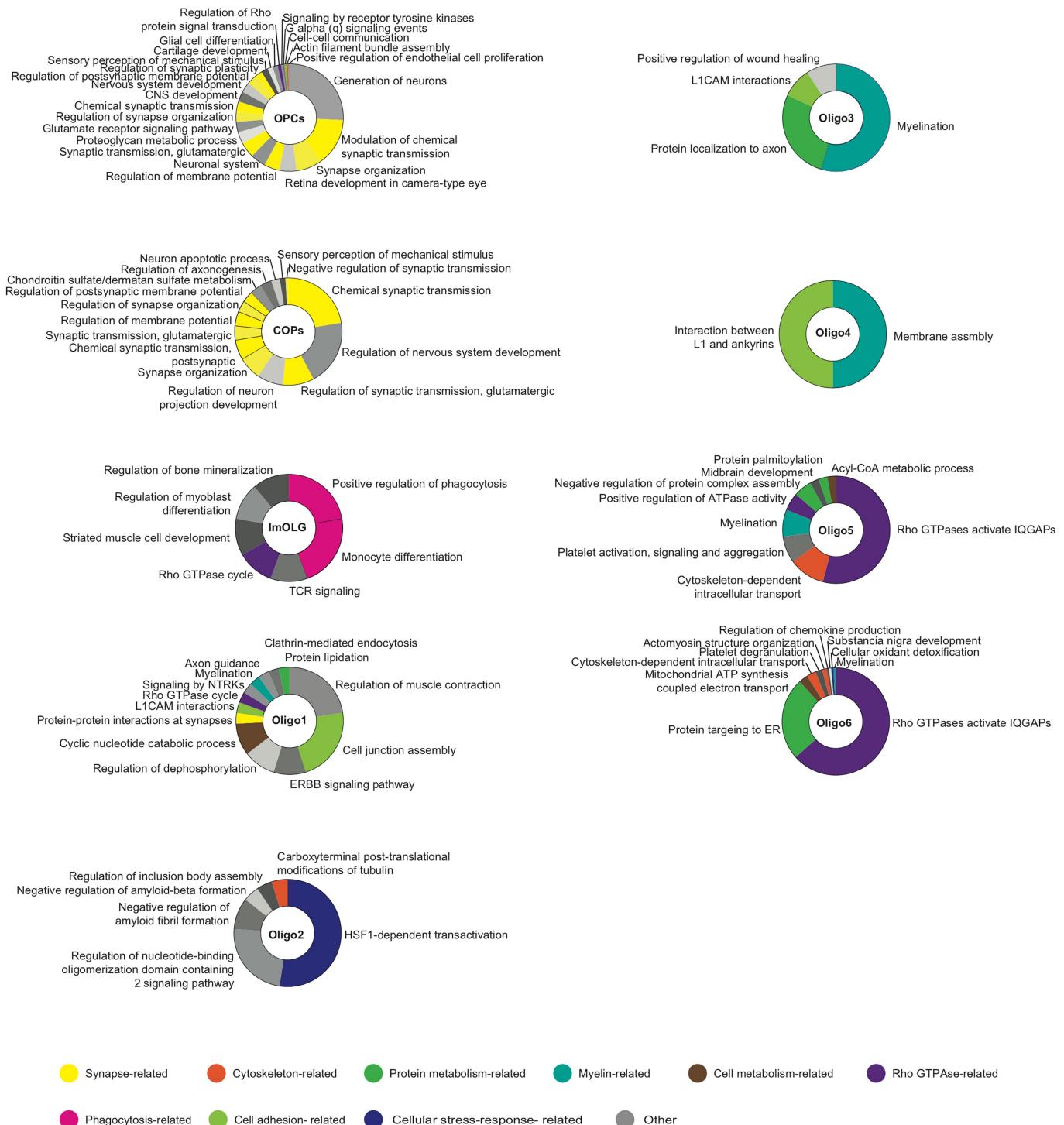
NOGO $^+$ OLIG2 $^+$ SOX6 $^-$ cells are mature oligodendrocytes. Scale bar, 10 μm . **f**, OPALIN staining of a control brain section. Scale bars, 5 mm and 300 μm (inset). **g**, OPALIN $^+$ Oligo6 nuclei in different bins of 300- μm increments from the grey–white matter border. Scale bar, 50 μm . $n = 3$ different control and MS individuals with NAWM and lesions.

P values determined by ANOVA. Data are mean \pm s.e.m. **h**, Combined OPALIN and KLK6 staining of another human control brain block. Scale bars, 5 mm and 50 μm (inset). In **b–e**, experiments were independently performed in two batches. **i**, Validation of oligodendrocyte mRNA markers in combination with OLIG1 and OLIG2 IHC. *BCAN* (top left), *CLDND1* (top right), *KLK6* (bottom left) and *CDH20* (bottom right). Red arrowheads denote marker $^+$ OLIG1/2 $^+$ oligodendrocytes; blue arrowheads denote marker $^-$ OLIG1/2 $^+$ oligodendrocytes. Scale bars, 10 μm .



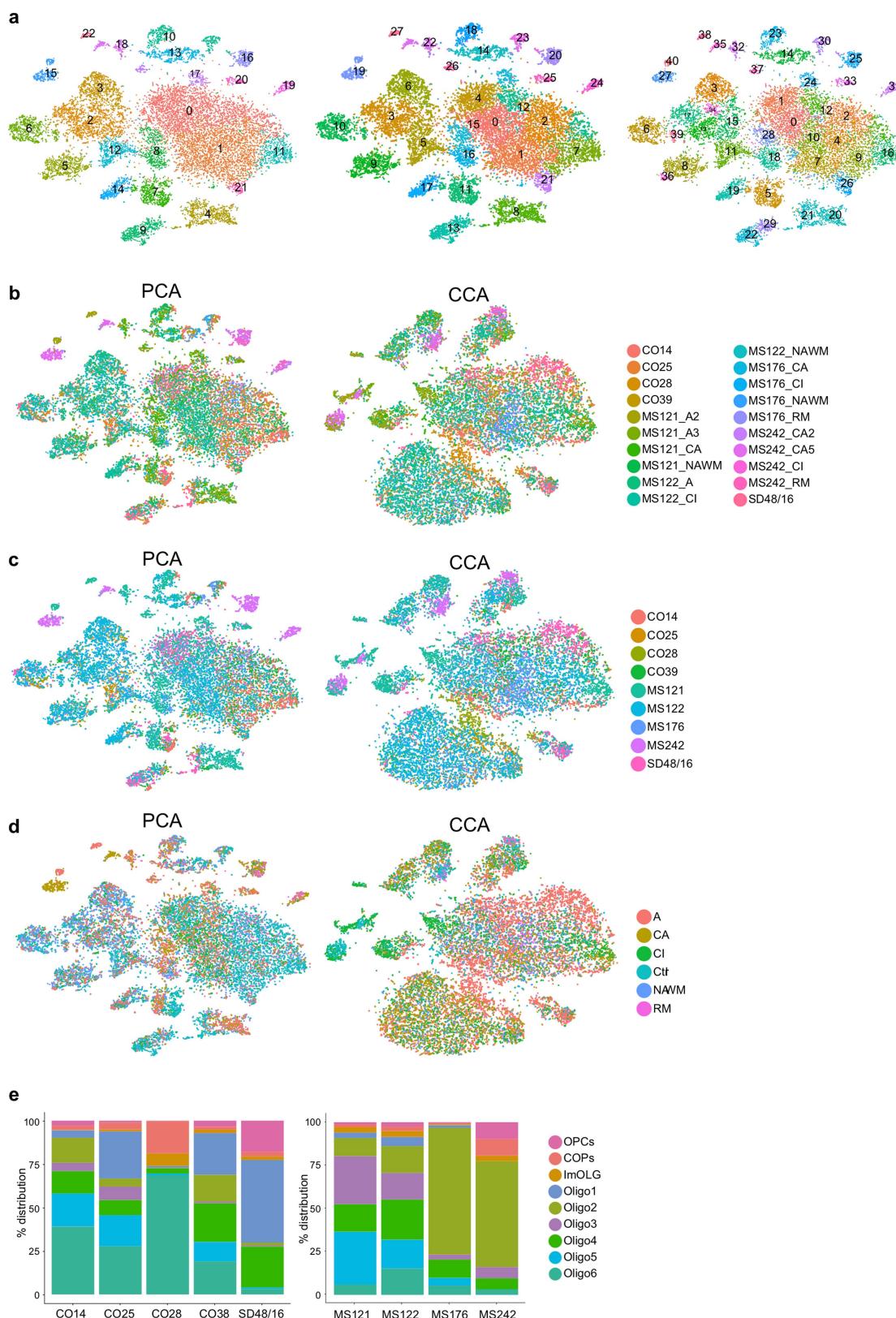
Extended Data Fig. 5 | Comparison of human control and MS oligodendrocyte snRNA-seq and mouse EAE oligodendroglia scRNA-seq datasets shows similarities and differences in oligodendrocyte

heterogeneity. Heat map of the mean AUROC values (see Methods) from the unsupervised classification of cell-type-to-cell-type comparisons between human (current dataset) and mouse⁴ oligodendroglia.



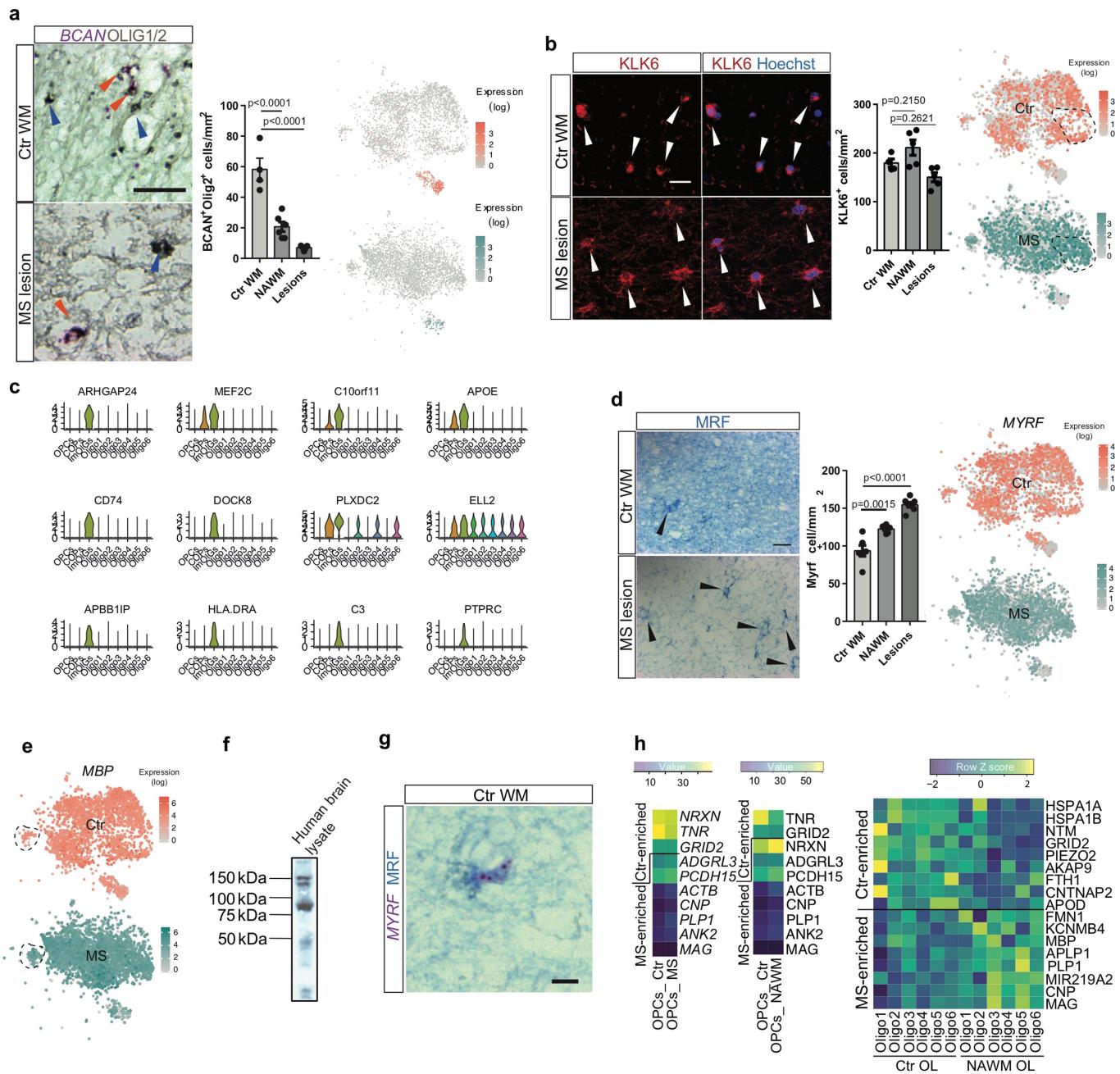
Extended Data Fig. 6 | GO analysis reveals functional differences between human oligodendrocyte subclusters. The most significantly differentially expressed genes from the snRNA-seq experiment of each oligodendrocyte subcluster were selected, and GO and pathway analysis

was performed with the ClueGO plug-in in Cytoscape on each individual cluster. Individual doughnut charts present the percentage of identified genes associated with the term and depict the most significant biological categories.



Extended Data Fig. 7 | Clustering of snRNA-seq dataset by different origins. **a**, t-SNE projections representing human control and MS white matter nuclei after dimensionality reduction with PCA at different resolutions. **b–d**, Clustering of snRNA-seq datasets by sample after dimensionality reduction with PCA (left) and canonical correlation

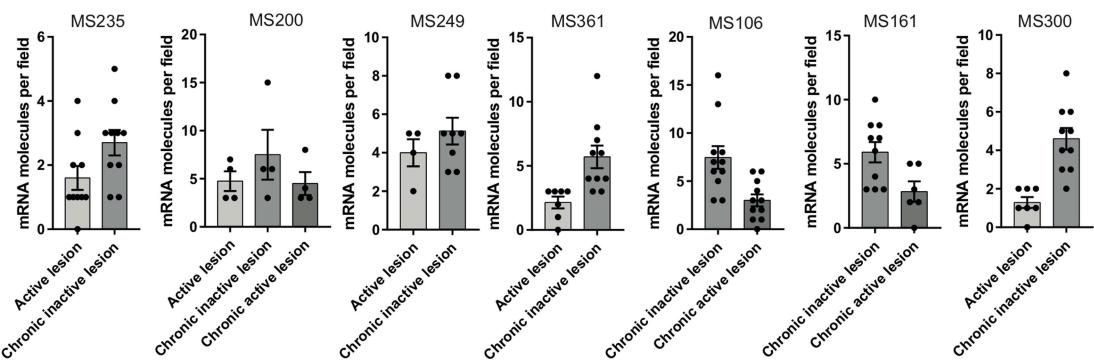
analysis (CCA) (right), highlighting control/MS individual and lesion type combined (**b**), control/MS individual (**c**) and lesion type (**d**) separately. **e**, Frequency distributions of oligodendrocyte subclusters by control (left) and MS (right) individuals. $n = 17,799$ cells derived from 5 control individuals and 4 patients with MS.



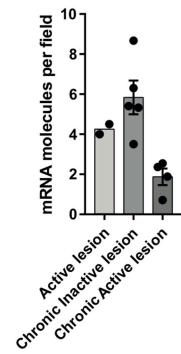
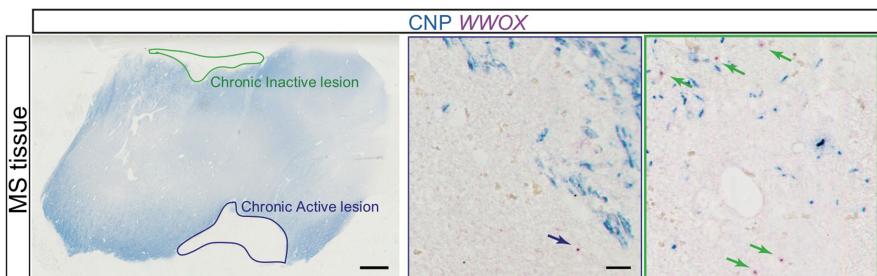
Extended Data Fig. 8 | Validation of skewed MS heterogeneity and oligodendrocyte gene expression profiling in control and NAWM tissues. **a**, Validation of BCAN-expressing OPCs in combination with OLIG1 and OLIG2 IHC. Red arrowhead denotes BCAN⁺OLIG1/2⁺ OPC; blue arrowhead denotes BCAN⁻OLIG1/2⁺ oligodendrocyte. Scale bar, 20 μm. t-SNE overlay of BCAN expression in the snRNA-seq dataset in control and MS samples. Scale bar, 20 μm. Data are mean ± s.e.m. $n = 4$ samples from different control individuals, $n = 6$ NAWM samples and $n = 5$ MS lesion samples from different patients with MS. P values determined by ANOVA. **b**, KLK6-expressing oligodendrocytes in control white matter, NAWM and MS lesions. Scale bar, 50 μm. Data are mean ± s.e.m. $n = 4$ samples from different control individuals; $n = 5$ different MS individuals. P values determined by ANOVA. t-SNE overlay of KLK6 expression in the control and MS snRNA-seq dataset. **c**, Violin plots showing the normalized expression counts of genes enriched in ImOLG in the snRNA-seq dataset (OPC $n = 352$; COP $n = 242$; ImOLG $n = 207$; Oligo1 $n = 1,129$; Oligo2 $n = 1,839$; Oligo3 $n = 775$; Oligo4 $n = 1,579$; Oligo5 $n = 1,167$; Oligo6 $n = 1,484$). Violin plots are centred

on the median with interquartile ranges, and the shape represents cell distribution. **d**, MRF IHC in control white matter, NAWM and MS lesions. Scale bar, 50 μm. Data are mean ± s.e.m. $n = 6$ samples from different control individuals, and $n = 7$ different patients with MS. P values determined by ANOVA. t-SNE overlay of MRF expression in the snRNA-seq dataset. **e**, t-SNE overlay of MBP expression in the control and the MS snRNA-seq dataset ($n = 4,037$ oligodendrocytes in control; $n = 4,737$ oligodendrocytes in MS). **f**, Western blot of the MRF antibody on human brain lysate to validate the specificity of the antibody. For gel source data, see Supplementary Fig. 1. **g**, Combination of MRF mRNA and protein labelling to confirm the specificity of the MRF antibody in control white matter. Scale bar, 10 μm. **h**, Heat maps representing the average gene expression of a subset of genes, including myelin-related genes, in control versus MS samples in OPCs (control versus MS and control versus NAWM) and mature oligodendrocytes (control versus NAWM). In **a**, **b** and **d**, each experiment was performed in two (three for **d**) independent batches, and P values are only displayed compared to control; in **f** and **g**, each experiment was performed twice on independent samples.

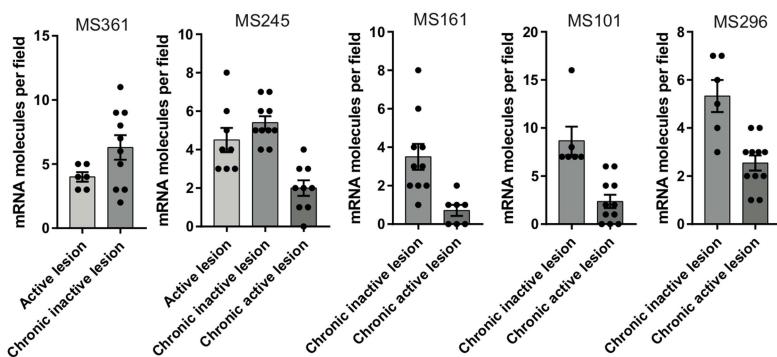
a CDH20



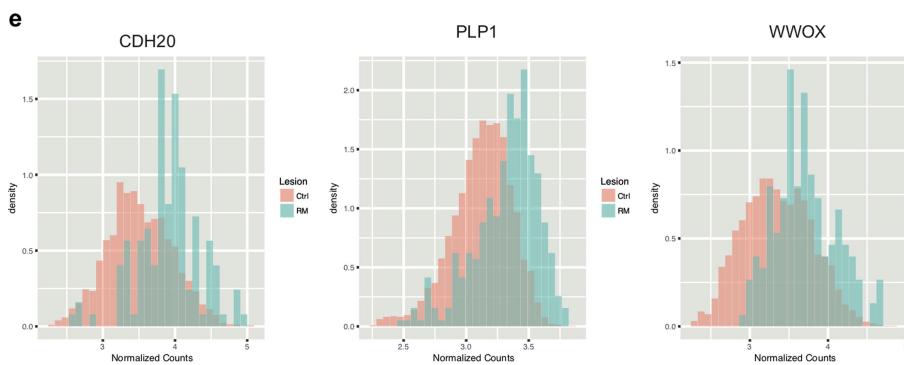
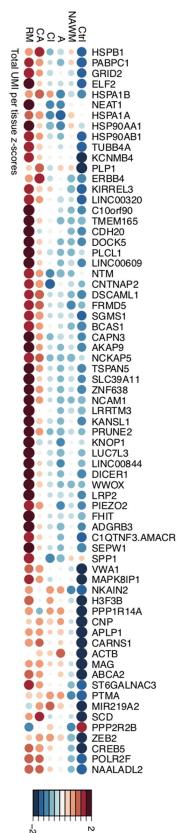
b



C WWOX



d



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Validations of altered oligodendrocyte heterogeneity in MS and mRNA expression differences in lesions.

a, Quantification of BaseScope *in situ* hybridization of *CDH20* in individual patients with MS (corresponds to Fig. 4c) shows an enrichment in chronic inactive lesions in each individual (n = individual number of quantified fields per patient ($n = 7$): MS235: $n = 10$ for active and chronic inactive lesions, MS200: $n = 4$ for active, and chronic inactive and chronic active lesions, MS249: $n = 4$ for active and $n = 8$ for chronic inactive lesions, MS361: $n = 7$ for active and $n = 10$ for chronic inactive lesions, MS106: $n = 11$ for chronic active and chronic inactive lesions, MS161: $n = 6$ for chronic active and $n = 10$ for chronic inactive lesions, MS300: $n = 7$ for active and $n = 10$ for chronic inactive lesions. Data are mean \pm s.e.m. **b, c**, BaseScope *in situ* hybridization of *WWOX* mRNA shows depletion of detected mRNA in chronic active lesions on average (**b**) and in individual patients with MS (**c**). Scale bars, 2 mm (left panel)

and 20 μm (middle/right panel). In **b**, $n = 2$ for active lesions and $n = 4$ for chronic inactive and chronic active lesions, in **c**, dots display the individual number of quantified fields per patient ($n = 5$), MS245: $n = 8$ for active, $n = 10$ for chronic inactive and $n = 9$ for chronic active lesions, MS361: $n = 6$ for active and $n = 10$ for chronic inactive lesions, MS101: $n = 6$ for chronic inactive and $n = 11$ for chronic active lesions, MS161: $n = 10$ for chronic inactive and $n = 7$ for chronic active lesions, MS296: $n = 11$ for chronic active and $n = 6$ for chronic inactive lesions. Data are mean \pm s.e.m. P values determined by ANOVA. **d**, Dot plot of the total normalized RNA UMI counts found within the lesions, NAWM and controls, in which both size and colour indicate z -scores. Blue and large denote low scores; red and large denote high scores; small denotes intermediate scores. **e**, Density histograms showing the difference in distribution of normalized counts observed between control and remyelinated lesions.

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 [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
- An indication of 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 statistics 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
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

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

Data collection

Zeiss Zen 2.3 Blue Edition
Leica Application SuiteX 3.5.2.18963
Vectra Polaris 1.0.7

Information regarding data collection is described in the methods section of the manuscript and data can be browsed at <https://ki.se/en/mbb/oligointernode>.

Data analysis

QuPath v0.1.2
Fiji-ImageJ v1.52h
Graph Pad Prism 7.
Cytoscape version 3.7.0

python packages : python 3.6
Cellranger version 2.1.1
Velocity.py 0.17.7

Own developed R code:
<https://github.com/Castelo-Branco-lab/GeneFocus>

R packages: R version 3.4.3 and R version 3.5:
 Velocyto.R_0.5
 Matrix_1.2-14
 Seurat_2.3.4 / Seurat.2.1.0
 ggplot2_3.0.0
<https://github.com/maggiecrow/MetaNeighbor.git>, 2017-08-28-runMN-US.R
 MAST_1.4.1
 SummarizedExperiment_1.8.1
 DelayedArray_0.4.1
 matrixStats_0.53.1
 GenomicRanges_1.30.3
 GenomeInfoDb_1.14.0
 S4Vectors_0.16.0
 Biobase_2.38.0
 BiocGenerics_0.24.0
 cowplot_0.9.2
 scater_1.8.1
 SingleCellExperiment_1.2.0

All tools described in the methods section of the manuscript, source code and notebooks are available at <https://github.com/Castelo-Branco-lab> under the repository Jaekel_Agirre_et_al_2018

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/reviewers upon request. 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

No restrictions apply on data availability. All figures have associated source code and/or raw data, these are: Fig 1-4 and Extended Data Figures 1-8. Sequence data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003412 . UMI expression and cell type annotation tables have been deposited in GEO, accession number GSE118257. The data can be explored and visualized in <https://ki.se/en/mbb/oligointernode>.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size

In the following section (also as displayed in the manuscript in the Figure legends and in the Materials and Methods section), n represents the individual number of different tissue donors (=different individuals), in case not stated differently. No sample size calculation was performed to pre-determine sample sizes. Rationale for sample size choice is provided for each of the experiments described below:

For the snRNA-seq experiment we used a sample size of n=5 Ctr and n=4 MS donors. We further sub-sampled the MS tissue into 3 normal appearing white matter (NAWM) samples from 3 different patients, 4 chronic active lesion samples from 4 different patients, 3 active lesion samples from 2 different patients, 3 chronic inactive lesion samples from 3 different patients and 2 remyelinating lesion samples from 2 different patients. These number of samples was sufficient to obtain the number of cells necessary to perform a confident data analysis. It was also possible to perform confident comparisons between controls and MS patients, between different lesion types and between this and already published datasets from human brain snRNA-seq experiments.

The chosen sample size for validations of findings arising from snRNA-Seq is in the same range of other publications in the research field using human post mortem donor tissue. The number of samples is limited by tissue availability from the brain banks. For the qualitative validation of mRNA markers in combination with immunohistochemistry, the sample size was a minimum of n=2. For the qualitative validation of double immunohistochemistry, the sample size was n=3.

In Fig.1c, we used n=4 for LURAP1L.-AS1+CDH20+, n=3 for the other combinations of OL subclass markers.

In Fig.3a we used n=5 Ctr and n=9 MS patients. In Fig.3b (SOX6) n represents the number of different donors. We used n=4 Ctr and n=5 NAWM and lesions.

In Fig. 3c (OPALIN) we used n=3 for Ctr and n=5 for NAWM and Lesions.
 In Fig.4e (CDH20 BaseScope) we used n=5 for active lesion, n=7 for chronic inactive lesions and n=3 for chronic active lesions in a total of 7 MS patients.
 For Extended Data Fig.4b-e we used n=3 different donors to validate the co-labeling of each marker.
 For Extended Data Fig.4g (OPALIN bins) we used n=3 different donors for each group.
 In Extended Data Fig. 8a (BCAN) we used n=4 Ctr, n=6 NAWM and n=5 Lesions. I
 In Extended Data Fig. 8b (KLK6) we used n=4 Ctr and n=5 NAWM and lesions.
 In Extended Data Fig.8d (MRF) we used n=6 Ctr and NAWM and n=7 lesions.
 For Extended Data Fig.9a, the individual number of quantified mRNA molecules per field per patient (n=7) are shown. We used the following number of fields: MS235: n=10 for A and CI lesions, MS200: n=4 for A, CI and CA lesions, MS249: n=4 for A and n=8 for CI lesions, MS361: n=7 for A and n=10 for CI lesions, MS106: n=11 for CA and CI lesions, MS161: n=6 for CA and n=10 for CI lesions, MS300: n=7 for A and n=10 for CI lesions.
 For Extended Data Fig.9b (WWOX) we used n=2 active lesions, n=5 chronic inactive lesions and n=4 chronic active lesions in a total of 5 MS patients. For Extended Fig.9c, the individual number of quantified mRNA molecules per field per patient (n=5) are shown. We used the following number of fields: MS245: n=8 for A, n=10 for CI and n=9 for CA lesions, MS361: n=6 for A and n=10 for CI lesions, MS101: n=6 for CI and n=11 for CA lesions, MS161: n=10 for CI and n=7 for CA lesions, MS296: n=11 for CA and n=6 for CI lesions.

Data exclusions	For single nuclei RNA-Seq data, we excluded data points through our quality control pipeline, as indicated in the methods section in the paper. In short, data was excluded based on thresholding of cells with transcript count less than 300, less than 200 expressing genes, more than 6000 UMI counts, a mitochondrial count ratio more than 0.20 and with 1 count above in at least 2 cells. Then, specific datasets from published studies were filtered specifically as explained in the methods section. After clustering we identify one nuclei cluster with high mitochondrial and ribosomal gene expression that was removed.
Replication	Verification of the experimental findings derived from single nuclei RNA-seq is performed through validation by mRNA (BaseScope) and protein (Immunohistochemistry) on a different set tissue sections derived from different donors. All validation data was reproducible and no patient datasets were excluded.
Randomization	For both the snRNA-seq and the in-situ validations, we used donor tissue from both sexes randomly distributed in each group (see manuscript Table 1). The age of the donors was also randomly chosen and distributed between the groups and does not display significant differences (see Table 1 and Methods section). For Immunohistochemistry, the fields for image quantification have been chosen randomly throughout the tissue section not to introduce any regional bias.
Blinding	The tissue for the snRNA-seq experiment could not be collected blinded, as specific areas had to be chosen. However, the experiment and subsequent data analysis has been performed fully automated to prevent experimenters' bias. Where possible, analysis of Immunohistochemistry has been performed automated with a image analysis software to exclude any bias. In the other cases, quantification has been validated by 2 independent people.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

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

Primary antibodies:

- anti-OLIG1 (Abcam, Cat. number ab68105, Lot number GR236765-4);
- anti-OLIG2 (rabbit, Atlas Antibodies, Cat. number HPA003254, Lot number CC81836);
- anti-OLIG2 (goat, R&D Systems, Cat. number AF2418, Lot number UPA0718031);
- anti-MYRF (Millipore, Cat. number ABN45, Lot number 2652011 and 2909641)
- anti OPALIN (Abcam, Cat number ab121425, Lot number GR264580-3)
- anti KLK6 (Life technologies, Cat number PA5-47239, TD2560533B)
- anti-CNP (Atlas Antibodies, Cat number AMAb91072, Lot number 02942)
- anti-SOX6 (Millipore, Cat number AB5805, Lot number 2921391)
- anti-NOGOA (R&D Systems, Cat number MAB3098, Lot number YYM0215021)

Secondary antibodies:

- rb-HRP IgG (Vector laboratories, Cat number MP-7401, Lot number ZD1019)
- rb-AP IgG (Vector laboratories, Cat number MP-5401, Lot number ZE0530)
- gt-HRP IgG (Vector laboratories, Cat number MP-7405, Lot number ZE0425)
- ms-HRP IgG (Vector laboratories, Cat number MP-7402, Lot number ZE0227)
- ms-AP IgG (Vector laboratories, Cat number MP-5402, Lot number ZC0831)

Colour/Fluorescent reaction kits:

- Fluorescein (Perkin Elmer, Cat number NEL741B001KT, Lot number 2779696)
- Cyanine 3 (Perkin Elmer, Cat number NEL744B001KT, Lot number 2328502)
- Cyanine 5 (Perkin Elmer, Cat number NEL745B001KT, Lot number 1859566)
- DAB (Vector laboratories, Cat number SK-4100, Lot number ZE0102)
- VectorBlue (Vector laboratories, Cat number SK-5300, Lot number ZE0402)

Validation

- anti-OLIG1: reactivity validated by the company for Rat, predicted to react with Human Mouse, Pig, Chimpanzee, Monkey, Baboon, Common marmoset. Has been used for human in Jakovcevski et al. 2005. Validated by the company for IHC on PFA-fixed frozen and fresh frozen tissue sections.
- anti-OLIG2: reactivity validated by the company for Human. Validated by the company for IHC. Has been validated by the Human Protein Atlas in 44 human control brain samples.
- anti-OLIG2: reactivity validated by the company for Human, Rat and Mouse. Validated by the company for IHC. Company has 30 citations for the use of this antibody.
- anti-MYRF: reactivity validated by the company for Human and Mouse. Validated by the company for IHC. MYRF antibody has further been validated by WB for specific binding as well as with a combination of mRNA and protein labeling of MYRF in the same cell (Extended Data Fig 8).
- anti OPALIN: reactivity validated by the company for Human. Validated by the company for IHC on paraffin-embedded tissue.
- anti KLK6: reactivity validated by the company for Human. Validated by the company for IHC on paraffin-embedded tissue.
- anti-CNP: reactivity validated by the company for Human, Rat and Mouse. Validated by the company for IHC and WB. Antibody has an entry at the Human Protein Atlas as validated antibody.
- anti-SOX6: reactivity validated by the company for Human and Mouse. Validated by the company for ICC.
- anti-NOGOA: reactivity validated by the company for Rat. This antibody has been used in human tissue in Kuhlmann et al. 2008. Validated by the company for IHC.

Human research participants

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

Population characteristics

For the snRNA-seq experiment:

The study included 5 Ctr (4 male, and 1 female) and 4 MS (3 male and 1 female) donors all in an age range between 35 and 82 years. Non of the Ctr donors had any known neurological disorders and no known medical treatment in this regard. The MS donors were all diagnosed with Chronic Multiple sclerosis (<14 years of disease duration) and non of them had a specific disease treatment.

For the in-situ validations:

The study included 11 Ctr (5 male and 6 female) and 15 MS (7 male and 8 female) donors in an age range between 36 and 77 years. Non of the Ctr donors had any known neurological disorders and no known medical treatment in this regard. The MS donors were all diagnosed with Chronic Secondary Progressive Multiple sclerosis (<18 years of disease duration for the cases with available information) and non of them had a specific disease treatment.

Detailed information is given in the Materials and Methods section of the manuscript and in Table1.

Recruitment

No donors were recruited, the tissue has been obtained from an accredited UK tissue bank.