Tools and Resources

Conservation and divergence of myelin proteome and oligodendrocyte transcriptome profiles between humans and mice

Vasiliki-Ilya Gargareta[1,#], Josefine Reuschenbach[1,#], Sophie B. Siems[1,#], Ting Sun[1,#,*], Lars Piepkorn[2,3], Carolina Mangana[1], Erik Späte[1], Sandra Goebbels[1], Inge Huitinga[4], Wiebke Möbius[1,5], Klaus-Armin Nave[1], Olaf Jahn[2,3,*] and Hauke B. Werner[1,*]

[1]Department of Neurogenetics, Max Planck Institute of Experimental Medicine, 37075 Göttingen, Germany
[2]Proteomics Group, Max Planck Institute of Experimental Medicine, 37075 Göttingen, Germany
[3]Translational Neuroproteomics Group, Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Georg-August-University, 37075 Göttingen, Germany
[4]Neuroimmunology Group, Netherlands Institute for Neuroscience, 1105 Amsterdam, The Netherlands
[5]Electron Microscopy Unit, Max Planck Institute of Experimental Medicine, 37075 Göttingen, Germany
[#]These authors contributed equally

***Correspondence**
Dr. Ting Sun, Department of Neurogenetics OR
Dr. Olaf Jahn, Proteomics Group OR
Dr. Hauke Werner, Department of Neurogenetics
Max Planck Institute of Experimental Medicine
Hermann Rein Str. 3
D-37075 Göttingen, Germany

Key words
Oligodendrocyte; myelin sheath; axon-glia interaction; label-free proteomics; scRNA-seq; cross-species comparison

Reviewer login information to the proteome data repository
ProteomeXchange Consortium PRIDE partner repository
Login data are provided in the Letter to the Editor

Word and Figure count
150 words (abstract), 4255 words (main text)
4 main figures, 8 figure supplements, 4 source data files

Abstract

Human myelin disorders are commonly studied in mouse models. Since both clades evolutionarily diverged approximately 85 million years ago, it is critical to know to what extent the myelin protein composition has remained similar. Here we use quantitative proteomics to analyze myelin purified from human white matter and find that the relative abundance of the structural myelin proteins PLP, MBP, CNP and SEPTIN8 correlates well with that in C57Bl/6N-mice. Conversely, multiple other proteins were identified exclusively or predominantly in human or mouse myelin. This is exemplified by peripheral myelin protein-2 (PMP2), which was specific to human CNS myelin, while tetraspanin-2 (TSPAN2) and connexin-29 (CX29/GJC3) were confined to mouse myelin. Assessing published scRNA-seq-datasets, human and mouse oligodendrocytes display well-correlating transcriptome profiles but divergent expression of distinct genes including *Pmp2, Tspan2* and *Gjc3*. Species-dependent diversity of oligodendroglial mRNA-expression and myelin protein composition can be informative when translating from mouse models to humans.

Introduction

Myelination of axons by oligodendrocytes facilitates rapid, saltatory impulse propagation in the central nervous system (CNS) of vertebrates [1]. The relevance of myelin for efficient motor, sensory and cognitive performance is illustrated by their decline in dysmyelinating and demyelinating disorders including multiple sclerosis (MS) and leukodystrophies and in respective mouse models [2]. MS is a human-specific autoimmune disorder for which it has remained difficult to establish a genuine mouse model, an observation that might point at the existence of human-specific antigens in myelin. Generated by mature oligodendrocytes (MOL), myelin consists of multiple concentric layers of specialized plasma membrane. The ultrastructure of myelin is highly ordered with closely apposed, compacted membrane layers and a non-compacted cytoplasmic channel system that includes the adaxonal myelin layer and paranodal loops. The formation of these sub-compartments is enabled by highly enriched, specialized myelin proteins. For example, the transmembrane-tetraspan proteolipid protein (PLP) supports extracellular membrane apposition and adhesion [3], the cytoplasmic myelin basic protein (MBP) mediates intracellular membrane apposition [4], and the enzyme cyclic nucleotide phosphodiesterase (CNP) contributes to structuring noncompact myelin compartments [5,6]. In fact, using the gel-based methods available at that time, PLP, MBP and CNP were early recognized as exceptionally abundant myelin proteins in the CNS of tetrapods [7,8]. Since then, the number of known myelin proteins has markedly increased, including proteins with enzymatic, cytoskeletal, adhesive and immune-related functions [1], and it became possible to quantify their relative abundance by mass spectrometry [9].

Myelin biology is primarily studied in mice and zebrafish [10]. The considerable differences between the species, including dimension and morphology of bodies and brains, motor performance, cognition and ecosystem, are owed to evolutionary changes since their last common ancestor about 420 million years ago (mya) [11]. Already early gel-based comparisons between CNS myelin fractions purified from various fish and tetrapod species revealed that the clades comprise overlapping but divergent sets of major myelin proteins [8,12]. More recently, this finding was extended to low-abundant constituents when quantitative mass spectrometry allowed comparing the CNS myelin proteome between zebrafish and mice [13]. By quantitative proteome analysis, MBP is highly abundant in CNS myelin of either species. Apart from MBP, however, their myelin proteome differs qualitatively and quantitatively. Thus, the protein composition of myelin displays species-dependent diversity, which can be assessed by mass spectrometry.

No animal model provides an exact replica of the human nervous system. The rodent and primate clades diverged approximately 85 mya [14]. In consequence, mice and humans went through considerable evolutionary time since their last common ancestor. To understand myelination and myelin-related diseases in humans it is thus relevant to investigate the molecular profiles of human oligodendrocytes and myelin, and, optimally, to compare them with their orthologs in relevant model species. There are evident ethical and methodological limitations to studies involving living humans. However, *post mortem* material donated for scientific assessment has become available. For example, the rates of oligodendrocyte turnover and myelin renewal have been evaluated in humans, including in multiple sclerosis patients [15]. It has also become possible to determine transcriptional profiles of oligodendrocytes in both humans and mice, including in disease conditions [16–18].

In the present study we used quantitative mass spectrometry to systematically examine the protein composition of myelin purified from the subcortical white matter of human subjects *post mortem*. Whereas the relative abundance of many structural myelin proteins is roughly similar between human and mouse CNS myelin – the latter as recently established by assessing c56Bl/6N mouse brains using the same methodology [9] - we observed striking qualitative and quantitative differences in the relative abundance of multiple other myelin proteins. By integrating and comparing previously established scRNA-seq datasets we found that their presence in myelin is largely reflected in the transcriptome profiles of mature oligodendrocytes (MOL). Our findings thus reveal unexpected differences in the molecular profiles of CNS myelin and oligodendrocytes between humans and mice. Considering their evolutionary divergence enables a more informed translation from mouse models to humans.

Results

Proteome analysis of human CNS myelin

To systematically identify and quantify the protein constituents of human CNS myelin, we biochemically purified a myelin-enriched light-weight membrane fraction from the normal-appearing white matter of five human subjects post-mortem. By electron microscopic assessment of the myelin fraction, constituents other than multilamellar myelin sheaths were largely absent (**Figure 1-supplement 1**), confirming that other membrane fractions had been efficiently removed.

We then subjected both the myelin fraction and the corresponding brain homogenate to solubilization using ASB-14 and high urea concentration, automated in-solution tryptic digest by filter-aided sample preparation (FASP), peptide fractionation by nanoUPLC, and ESI-QTOF mass spectrometry involving data-independent acquisition (DIA) of data. The utilized MS^E mode facilitates simultaneous quantification and identification of all peptides entering the mass spectrometer. Proteins can thus be quantified by correlating signal intensities of peptides with those of a spike-in protein of known concentration (TOP3 method; [19]). When assessing myelin by MS^E we quantified 332 proteins (**Figure 1-source data 1**; labeled in green in **Figure 1a**) with a false discovery rate (FDR) of <1% and an average sequence coverage of 39.6%. When using the ultra-definition (UD)-MS^E mode, in which ion mobility spectrometry enables an additional separation of peptides after chromatography and before mass measurement, we identified and quantified 835 proteins with an average sequence coverage of 37.0% (**Figure 1-source data 1**; labeled in blue in **Figure 1a**). The MS^E mode quantified myelin proteins with a dynamic range of over four orders of magnitude parts per million (ppm), thereby allowing the reliable quantification of all myelin constituents including the exceptionally abundant PLP, MBP and CNP. The $UDMS^E$ mode identified over twice as many proteins as MS^E, though with a compressed dynamic range of only about three orders of magnitude ppm. Expectedly, the MS^E and $UDMS^E$ datasets correlated well with a correlation coefficient of >0.8 (**Figure 1b; Figure 1-supplement 2**). Both datasets taken together, we identified 848 proteins in human CNS myelin by LC-MS analysis. Importantly, the strategy of direct label-free quantification provides information about the relative abundance of identified proteins. When comparing the relative abundance of proteins in the myelin fraction and the corresponding homogenate, we found known myelin markers enriched in the myelin fraction (**Figure 1c**). Markers for other cell types or compartments were either reduced in abundance in the myelin fraction compared to brain lysate or not identified at all (**Figure 1-supplement 3**). This indicates that the fraction is suited for proteomic analysis of human myelin.

158 **Relative abundance of CNS myelin proteins in humans**

159 We used the MS^E dataset to calculate the relative abundance of myelin proteins in the human
160 white matter (**Figure 1d**), considering that quantification of exceptionally abundant proteins
161 requires a high dynamic range. The most abundant myelin proteins were the structural
162 constituents proteolipid protein (PLP), myelin basic protein (MBP) and cyclic nucleotide
163 phosphodiesterase (CNP), which accounted for 44.8%, 28.4% and 4.5% of the total myelin
164 protein, respectively. In addition, numerous known myelin proteins were identified and
165 quantified at lower abundance (**Figure 1d**). Previously known myelin proteins constituted
166 approximately 82% of the total human myelin protein (**Figure 1d**), while the remaining 18%
167 were accounted for by other proteins, including occasional contaminants from other cellular
168 sources (**Figure 1-supplement 3**).
169
170 **Comparison to the mouse myelin proteome**
171 We hypothesized that the protein composition of human and mouse myelin displays some
172 degree of divergence. To compare human and mouse myelin, we first separated myelin of both
173 species by SDS-PAGE. By silver staining, the band patterns were roughly comparable but not
174 identical (**Figure 2a**), supporting the hypothesis that some differences exist. To elucidate
175 differences at the molecular level, we compared the present human mass spectrometric data
176 with those of our recent proteomic analysis of myelin purified from the brains of C57Bl/6N mice
177 using the same workflow and methodology [9] (ProteomeXchange Consortium PRIDE partner
178 repository, dataset identifier PXD020007). As expected, the majority of known myelin proteins
179 was identified in myelin of both species (**Figure 2b**). However, a subset of known myelin
180 proteins was identified only in either human or mouse myelin (**Figure 2b**), in agreement with
181 the hypothesis that the protein composition of myelin is not identical across these species.
182
183 For example, we noted that peripheral myelin protein 2 (PMP2, also termed P2 or fatty acid
184 binding protein (FABP8)) was identified in human CNS myelin (**Figure 2b**). PMP2 has long
185 been known as a constituent of myelin in the peripheral nervous system (PNS) synthesized by
186 Schwann cells [20,21] but based on rodent studies was assumed to be absent from CNS myelin.
187 Yet, PMP2 was readily detected in human CNS myelin by both immunoblotting (**Figure 2-**
188 **supplement 1a**) and immunohistochemistry (**Figure 2-supplement 1b**), thus confirming its
189 mass spectrometric identification. In contrast, PMP2 was not detected in mouse CNS myelin
190 by immunoblot (**Figure 2-supplement 1a**). In agreement with prior work [20] PMP2 was readily
191 detected in mouse PNS myelin (**Figure 2-supplement 1a**), indicating that the utilized
192 antibodies detect PMP2 of either species. Together this substantiates the existence of species-
193 dependent differences in the protein composition of CNS myelin between humans and mice.
194

Next we plotted all proteins identified in human CNS myelin, i.e. the present MS^E and $UDMS^E$ datasets, against those identified in mouse myelin as recently established using the same methodology[9] (ProteomeXchange Consortium PRIDE partner repository, dataset identifier PXD020007). Indeed, the datasets correlated well with correlation coefficients of 0.88 (MS^E) and 0.59 ($UDMS^E$) (**Figure 2c,d**) but clearly diverged to some extent. We therefore cross-compared the abundance of individual myelin proteins in human and mouse myelin by MS^E using heatmap visualization (**Figure 2e**). We found that major structural myelin proteins including PLP, MBP, CNP, SEPTIN2, SEPTIN7 and SEPTIN8 displayed a similar relative abundance in myelin of both species. However, several other myelin proteins were comparatively more abundant in human myelin, as exemplified by crystallin-αB (CRYAB), CD9 (also termed tetraspanin-29/TSPAN29), and peptidyl arginine deiminase (PADI2), or in mouse myelin, including myelin-associated oligodendrocyte basic protein (MOBP), sirtuin-2 (SIRT2), and carbonic anhydrase 2 (CA2). Importantly, when detecting these proteins by immunoblotting in myelin of both species (**Figure 2f**) these results were generally consistent with the mass spectrometric comparison (**Figure 2e**). Yet, quantitative mass spectrometry emerged as more straightforward than immunoblotting when comparing the relative abundance of proteins across species if species-dependent differences in splice isoforms exist. This is exemplified by MBP, which - owing to species-dependent alternative splicing [22] - displays three main isoforms (14.0, 17.0 and 18.5 kDa) in mouse CNS myelin but only one dominant isoform (18.5 kDa) in human CNS myelin, in agreement with previous observations [23,24]. Taken together, the protein composition of human and mouse CNS myelin is similar with respect to the relative abundance of major structural proteins but displays remarkable qualitative and quantitative differences regarding many other myelin proteins.

Integrated scRNA-seq profile of human and mouse mature oligodendrocytes (MOL)

To identify species-dependent transcriptional differences that may underlie the diversity of the myelin proteome, we utilized high-resolution mRNA-abundance profiles to assess the oligodendrocyte lineage in both humans and mice. To this aim, we retrieved previously published scRNA-seq datasets from the CNS of humans [17,18,25–28] and mice [16,28–33] and evaluated all cells designated as oligodendrocyte progenitor cells (termed OPC in the following), newly formed oligodendrocytes (termed NFO) and mature oligodendrocytes (termed MOL) from non-diseased subjects (**Figure 3-supplement 1a,b, Figure 3-source data 1)**. Using the SCTransform pipeline within the R toolkit Seurat, it was possible to integrate cells from all available human and mouse datasets into respective single objects (**Figure 3-supplement 1a,b**). Importantly, cells from all studies distributed well across the uniform manifold approximation and projection (UMAP)-plots (**Figure 3-supplement 1a,b**), implying suitability for integration and further assessment. Indeed, when highlighting marker gene

expression on UMAP projections, cells expressing markers for OPCs (*CSPG4, PCDH15, PDGFRa, PTPRZ1*) or MOL (*ANLN, CNP, MBP, PLP1*) clustered well in both the human and mouse integrated datasets (**Figure 3-supplement 1c,e**). Notably, however, multiple myelin-related transcripts displayed considerable expression only in human or mouse oligodendrocytes, as exemplified by *TSPAN2*, *GJC3* and *PMP2* (**Figure 3-supplement 1f**). We noted that cells expressing NFO markers (*BMP4, ENPP6, FYN, GPR17*) clustered well in the mouse but not the human integrated dataset (**Figure 3-supplement 1d**), probably owing to the low number of NFO in the latter. Indeed, only 132 NFO were comprised in the human scRNA-seq datasets, considerably fewer compared to 10,391 NFO recovered from the mouse datasets (**Figure 3-source data 1)**. For this reason we focused on mature oligodendrocytes (MOL) for a more thorough species-dependent comparison of transcriptional profiles of myelin-related genes.

To this aim we subset all cells annotated as MOL from control samples in all datasets of both species [16–18,25–33] (**Figure 3a**) for integration via SCTransform. Mouse gene symbols were translated to human gene symbols prior to data integration. Importantly, cells from both species distributed well across the UMAPs (**Figure 3b,c**), providing the basis for assessing the transcriptional profiles of 41,517 human and 95,966 mouse MOL. At the level of gene expression, cells expressing myelin marker transcripts (*ANLN, CNP, GSN, MBP, PLLP, PLP1*) distributed similar across human and mouse MOL (**Figure 3d**), as did transcripts encoding myelination-related transcription factors (*MYRF, SOX10*) (**Figure 3e**). Notably, multiple myelin-related transcripts displayed exclusive or predominant expression in MOL of only one of the species, as exemplified by *PMP2, PADI2, CA2, TSPAN2 and GJC3* (**Figure 3f**). Also, when assessed in the species-separately integrated datasets of all oligodendroglial cells including OPC, NFO and MOL, these genes displayed little or no expression in the respective other species (**Figure 3-supplement 1f**).

To compare the transcriptome profiles of human and mouse MOL without the influence of sequencing batch effects, we applied van der Waerden (vdW)-ordered quantile transformation to the mean mRNA abundance values of 3,000 integration features (i.e., genes) in all cells designated as MOL in all individual datasets and calculated the average relative transcript abundance in both human and mouse MOL. The averaged vdW-normalized mRNA-abundance profiles correlated reasonably well with a coefficient of 0.59 (**Figure 3-supplement 2**). However, a higher correlation coefficient of 0.82 was found when comparing only known myelin-related transcripts (n=34 transcripts; highlighted as blue data points in **Figure 3-supplement 2**). This indicates that the transcriptional profiles of known myelin-related transcripts are more similar between humans and mice than those of other transcripts

expressed in MOL. In particular, the abundance of several transcripts encoding structural myelin proteins, including *PLP1*, *MBP* and *CNP,* was essentially equal between human and mouse MOL. Also the abundance of transcripts for myelin-related transcription factors (*MYRF*, *OLIG1*, *OLIG2*, *SOX10*) was roughly similar (labelled in green in **Figure 3-supplement 2**). We noted that the most abundant transcripts in MOL also included genes of which the protein products were not mass spectrometrically identified in myelin - and thus not comprised in the myelin proteome - either because they are secreted (*FTH1*, *TF*, *APOD*) or display a non-suitable tryptic digest pattern (MAL) (labelled in orange **Figure 3-supplement 2**). The abundance of these transcripts also correlated well between human and mouse MOL. As an exception, the abundance of *APOE* mRNA was considerably higher in mouse MOL compared to human MOL (**Figure 3-supplement 2**).

Taken together, the integrated scRNA-seq profiles of MOL generally correlated well between humans and mice, in particular with respect to most known myelin-related mRNAs. However, multiple distinct transcripts were found with qualitatively or quantitatively divergent abundance when compared between the species. It is also noteworthy that the degree of correlation is lower when comparing the myelin proteome (by either MS^E or $UDMS^E$) with the averaged vdW-normalized mRNA abundance profile of MOL in both humans and mice (**Figure 3-supplement 3**).

Subpopulation analysis of integrated human and mouse MOL scRNA-seq profile

Previously, multiple transcriptome studies have identified distinct subpopulations of MOL in both humans and mice [16,17,29], which were correlated according to the expression of marker genes. Here, we tested whether similar subpopulations also manifest if evaluating the merged and integrated scRNAseq profiles, thereby not only allowing cross-species comparison but also increasing the dimensionality of assessed MOL per species. Indeed, k-nearest neighbor (KNN) clustering identified five potential subpopulations of MOL (labelled as clusters 0, 1, 2, 3 and 4 in **Figure 4a-c**). Notably, all subpopulations displayed approximately similar expression levels of marker genes encoding structural myelin proteins (*PLP1, MBP, CNP, CLDN11, MAG*) (**Figure 4b**). However, the subpopulations were defined by varying degrees of expression of other transcripts, including the myelin-related *CD9* and *OPALIN* (cluster 0), *APOD, KLK6* and *S100B* (cluster 1), *APOE* and *CST3* (cluster 2), *CA2* and *PTGS* (cluster 3), and *SIRT2* and *NFASC* (cluster 4) (**Figure 4b**). Considering the larger number of evaluated cells compared to the prior individual studies on which the present assessment is based, these findings support the previously identified subpopulations of MOL[17,29]. Based on gene ontology (GO) term enrichment analysis of biological processes (**Figure 4-supplement 1**), one may speculate that MOL in clusters 2, 3 and 4 are associated with GO terms grouped as protein synthesis, electron

306 transport and immune activation, respectively. However, their functional specialization and
307 relevance remain to be shown. Less speculatively, both human and mouse MOL comprise all
308 five subpopulations to an approximately similar extent (**Figure 4c**), implying that none of these
309 MOL subpopulations is restricted to either one of these species.
310
311

Discussion

We performed quantitative proteome analysis to determine the protein composition of human CNS myelin. Subjecting myelin biochemically purified from human subcortical white matter to label-free mass spectrometry allowed identifying hundreds of proteins with very high confidence. More importantly, the method involves quantifying peptide intensities without prefractionation, thereby providing direct information about the relative abundance of myelin proteins. The latter provides a considerable advancement compared to previous approaches involving prefractionation at the protein level via 1D-gels [24,34] or at the peptide level via 2D-liquid chromatography [35,36], which yielded lists of proteins identified in human CNS myelin but without information about their relative abundance.

Knowing the relative abundance of myelin proteins enables both considering their stoichiometric relationships and cross-species comparisons. For example, the filament-forming septins SEPTIN2, SEPTIN4, SEPTIN7 and SEPTIN8 displayed a molar stoichiometry of about 1:1:2:2 in human CNS myelin. Notably, the same septin subunits are also comprised in myelin of mice with a similar molar stoichiometry [9], a likely prerequisite for their assembly into similar core multimers and higher order structures. Indeed, experiments in mice have previously shown that these septin subunits assemble into membrane-associated filaments that stabilize the adaxonal compartment of CNS myelin [37]. Integrating the current view on septin assembly [38] and the relative abundance of septin subunits in myelin it is possible to deduce that the predominant core multimer in myelin is a hexamer of septins 2/4-8-7-7-8-2/4. The comparatively low abundance of SEPTIN9 in myelin implies that core octamers occur less frequently. Assessing the relative abundance of myelin proteins also allows deducing that for each core hexamer up to one molecule of the adaptor protein anillin that facilitates septin assembly [39] is present in CNS myelin. Together, the relative abundance and multimer composition of myelin septins emerges as conserved between human and mouse CNS myelin, similar to that of other structural myelin proteins, including PLP, MBP and CNP.

On the other hand, we also found considerable qualitative and quantitative differences between the protein composition of human and mouse CNS myelin. For example, the tetraspan-transmembrane proteins TSPAN2 [40–42] and GPM6B/M6B/Rhombex29 [43,44] were previously established as myelin proteins in mice and rats, and they were readily identified in CNS myelin of mice by both mass spectrometry [9] and immunoblot. However, these proteins were of very low abundance or virtually undetectable in human myelin by both techniques. It has been established in experimental mice that TSPAN2 and GPM6B contribute to immunomodulation [42] and myelin biogenesis [44], respectively. The present data thus imply that

the mouse and human orthologs do not contribute equally to these functions. The gap junction protein GJC3/CX29 has also been established as a myelin protein in mice and rats and is thought to mediate intercellular coupling via gap junctions [45,46]. Similar to TSPAN2 and GPM6B, GJC3 has been readily identified mass spectrometrically and by immunoblotting in mouse but not human myelin. Interestingly, though, deletion of the *Gjc3*-gene in mice did not have evident morphological or functional consequences for CNS myelin or oligodendrocytes [47,48]. The benefit for mouse CNS myelin to comprise GJC3 thus remains unknown at this time.

On the other hand, peripheral myelin protein 2 (PMP2, previously termed P2 or FABP8), a membrane-phosphoinositide-binding protein [49], has long been known as a constituent of peripheral myelin generated by Schwann cells in the PNS [21,50,51] and actually considered a marker to discriminate peripheral from central myelin [8,52,53]. Notably, the experiments establishing this view involved bovine and rodent but not human samples. Our finding that PMP2 is a myelin protein in the human CNS leaves open the question of whether this reflects clade-specific *de novo* recruitment into CNS myelin or selective constraints that eliminated PMP2 from myelin in the clade including cows and rodents. At the evolutionary level, it is interesting to speculate which benefits (if any) human myelin may have from comprising PMP2, or what the evolutionary constraints may be that led to its dropout from rodent CNS myelin. A lead may come from the investigation of the PNS of *Pmp2*-deficient mice, which displayed an altered myelin lipid profile associated with reduced motor nerve conduction velocity [54]. It is tempting to speculate that presence of PMP2 in human but not mouse CNS myelin may affect the composition or organization of its lipids, and, possibly, conduction velocity. The *PMP2* gene causes – when mutated – the peripheral neuropathy Charcot-Marie-Tooth (CMT) disease type 1G [55,56]. A subset of these patients has been tested by brain MRI; however, no major pathology of the white matter was found that would be typical of a leukodystrophy [55]. Yet, our finding that PMP2 is a myelin protein in the human CNS indicates that further testing these and other CMT1G patients for central involvement may find yet-overlooked impairments, possibly more subtle than visible by MRI.

How may species-dependent differences in myelin protein composition come about at the molecular level? Considering the limitations imposed by the availability of human samples, we can not formally rule out that differences in the sex or age of specimen, brain region, sample preparation or data analysis may affect the degree of correlation. However, we note that both male and female donors are represented in the human samples, and that machinery, methodology and data analysis were the same in establishing the mouse myelin proteome [9] and the human myelin proteome assessed here. The *post mortem*-delay unavoidable for sampling human specimen is unlikely to affect the present comparison when considering that

386 the average *post mortem*-delay is six hours at the Netherlands brain bank that supplied the
387 human samples used here and that an experimental *post mortem*-delay of six hours did not
388 considerably affect the myelin proteome in C57Bl/6N mice [9]. Finally, we believe that the high
389 degree of cross-species similarity regarding the abundance of structural myelin proteins
390 between humans and mice allows trust in the overall comparison of myelin protein
391 composition, including for proteins displaying cross-species dissimilarity. Thus, individual
392 myelin proteins displaying species-dependent differences may be owing to species-dependent
393 differences in intracellular trafficking and incorporation into the myelin sheath, stability and
394 turnover rate, mRNA-to-protein translation efficiency, or actual mRNA expression.
395
396 Indeed, our cross-species integration and comparison of the scRNAseq profiles of mature
397 oligodendrocytes (MOL) implies that species-dependent mRNA expression can explain the
398 species-dependent differences in myelin protein composition at least to some extent. For
399 example, *PMP2/Pmp2* mRNA is expressed in human but not mouse MOL and PMP2 protein
400 was identified in human but not mouse myelin. *Vice versa*, *TSPAN2/Tspan2* and *GJC3/Gjc3*
401 transcripts are expressed in mouse but not human MOL and their protein products TSPAN2
402 and GJC3 are identified in mouse but not human myelin. Less exclusively, *CA2/Car2* mRNA
403 is preferentially expressed in mouse compared to human MOL, correlating with the relative
404 abundance of its protein product CAR2/CA2 in mouse compared to human myelin. *Vice versa*,
405 the higher abundance of PADI2 in human compared to mouse myelin goes along with a higher
406 abundance of *PADI2/Padi2* mRNA in human compared to mouse MOL. Together, species-
407 dependent protein abundance in myelin is probably owing to species-dependent mRNA
408 expression, at least for some myelin constituents. We speculate that the evolutionary
409 emergence of regulatory elements that regulate oligodendroglial gene expression in the
410 hominin clade [57] partly underlies speciation of oligodendroglial transcript profiles and myelin
411 protein composition. However, evolutionary changes may also affect oligodendroglial gene
412 regulation specifically in the rodent clade.
413
414 It is of note that not all myelin constituents display an evident cross-species correlation
415 between the abundance of its transcript in myelinating oligodendrocytes and the abundance
416 of the protein product in myelin. Examples include the tetraspanin CD9/TSPAN29 [41,58], which
417 is more abundant in human compared to mouse myelin, and the enzyme aspartoacylase
418 (ASPA) [59], which is more abundant in mouse compared to human myelin. Indeed, the
419 abundance of the transcripts encoding either protein are approximately similar when
420 comparing human and mouse MOL. This indicates that not all differences in the protein
421 composition of human and mouse myelin are caused by species-dependent differences in
422 gene expression by MOL. Instead, the efficiency of mRNA translation, intracellular trafficking,

423 incorporation into myelin, stability or turnover rate of myelin proteins may display species-
424 dependent efficiency.
425
426 In conclusion, both oligodendroglial mRNA abundance profiles and the CNS myelin proteome
427 display widespread similarities between humans and mice, suggesting considerable
428 evolutionary conservation. However, distinct molecular differences were evident, indicating
429 evolutionary recruitment or dropout of myelin proteins across mammalian clades. Mice are
430 commonly assessed as a model for humans in myelin biology. Considering the evolutionary
431 heterogeneity of oligodendroglial mRNA expression and myelin composition can be instructive
432 when translating between mouse models and humans.
433
434

Methods

Human samples

Post-mortem brain tissue was provided by the Netherlands Brain Bank. Donors gave informed consent to perform autopsy and for the use of clinical and pathological information by researchers, approved by the medical ethics committee of the VU medical center (Amsterdam, The Netherlands). The diagnoses were confirmed by a neuropathologist.

The following subjects were used for myelin purification and proteome analysis of lysate (L) and purified myelin (M) as well as for immunoblotting.

Subject	Sex	Age (years)	Diagnosis	PMD (hours)	Brain region	Sample name in mass spectrometry
1995-106	male	74	Non-demented control	08:00	White matter	Control_L1/2 Control_M1/2
1996-052	male	73	Non-demented control	09:10	White matter	Control_L3/4 Control_M3/4
2002-024	female	75	Non-demented control	05:30	White matter	Control_L5/6 Control_M5/6
2010-015	female	73	Non-demented control	07:45	White matter	Control_L7/8 Control_M7/8
2017-124	female	55	Non-demented control	07:30	White matter	Control_L9/10 Control_M9/10

The following subjects were used for immunohistochemical analysis:

Subject	Sex	Age (years)	Diagnosis	PMD (hours)	Brain Region	Comments
2019-026	male	55	Parkinson Disease	05:55	Optic nerve	No pathology of the optic nerve
2019-077	female	91	Non-demented control	09:30	Optic nerve	Immunolabeled section shown in **Figure 2-supplement 1** originates from this subject
2019-106	female	80	Non-demented control	06:50	Optic nerve	-

Animal welfare

For the procedure of sacrificing vertebrates for preparation of tissue, all regulations given in the German animal welfare law (TierSchG §4) are followed. Since sacrificing of vertebrates is not an experiment on animals according to §7 Abs. 2 Satz 3 TierSchG, no specific ethical review and approval or notification is required for the present work. All procedures were

supervised by the animal welfare officer and the animal welfare committee for the Max Planck Institute of Experimental Medicine, Göttingen, Germany. The animal facility at the Max Planck Institute of Experimental Medicine is registered according to §11 Abs. 1 TierSchG.

Myelin purification

A light-weight membrane fraction enriched for myelin was purified from pieces of normal appearing white matter of human subjects post-mortem as specified above, brains of C57Bl/6N mice, and sciatic nerves of C57Bl/6N mice using an established protocol involving two steps of sucrose density gradient centrifugation and osmotic shocks [60]. Myelin accumulates at the interface between 0.32 M and 0.85 M sucrose.

Electron microscopy of purified myelin

For assessment of the human myelin fraction by electron microscopy, myelin purified from the white matter of subjects 1995-106 and 1996-052 was used. 75µl of each myelin sample was mixed with 75µl 2x concentrated fixative composed of 5% glutaraldehyde, 8% formaldehyde and 1.0% NaCl in 100 mM phosphate buffer pH 7.3. Then the fixed fraction was spun down and resuspended in 2% agarose Super LM (Roth, Karlsruhe, Germany). After solidification the pellet was cut into two halves and embedded in Epon after postfixation in 2% OsO_4. Ultrathin sections across the pellet were prepared using an UC7 ultramicrotome (Leica Microsytems, Vienna, Austria) equipped with a 35° diamond knife (Diatome, Biel, Switzerland). Images were taken with a LEO912 transmission electron microscope (Carl Zeiss Microscopy, Oberkochen, Germany) using a 2k on-axis CCD camera (TRS, Moorenweis, Germany).

Label-free quantification of myelin proteins

In-solution digestion of myelin proteins according to an automated filter-aided sample preparation (FASP) protocol [60] and LC-MS-analysis by different MS^E-type data-independent acquisition (DIA) mass spectrometry approaches was performed as recently established for mouse PNS [61] and CNS [9] myelin. Briefly, protein fractions corresponding to 10 µg myelin protein were dissolved in lysis buffer (1% ASB-14, 7 M urea, 2 M thiourea, 10 mM DTT, 0.1 M Tris pH 8.5) and processed according to a CHAPS-based FASP protocol in centrifugal filter units (30 kDa MWCO, Merck Millipore). After removal of the detergents, protein alkylation with iodoacetamide, and buffer exchange to digestion buffer (50 mM ammonium bicarbonate (ABC), 10 % acetonitrile), proteins were digested overnight at 37°C with 400 ng trypsin. Tryptic peptides were recovered by centrifugation and extracted with 40 µl of 50 mM ABC and 40 µl of 1% trifluoroacetic acid (TFA), respectively. Combined flow-throughs were directly subjected to LC-MS-analysis. For quantification according to the TOP3 approach [19], aliquots were spiked

with 10 fmol/µl of Hi3 EColi standard (Waters Corporation), containing a set of quantified synthetic peptides derived from the *E. coli* chaperone protein ClpB.

Nanoscale reversed-phase UPLC separation of tryptic peptides was performed with a nanoAcquity UPLC system equipped with a Symmetry C18 5 µm, 180 µm × 20 mm trap column and a HSS T3 C18 1.8 µm, 75 µm × 250 mm analytical column (Waters Corporation) maintained at 45°C. Peptides were separated over 120 min at a flow rate of 300 nl/min with a gradient comprising two linear steps of 3-35% mobile phase B (acetonitrile containing 0.1% formic acid) in 105 min and 35-60% mobile phase B in 15 min, respectively. Mass spectrometric analysis on a quadrupole time-of-flight mass spectrometer with ion mobility option (Synapt G2-S, Waters Corporation) was performed in the ion mobility-enhanced DIA mode with drift time-specific collision energies referred to as UDMS^{E} [62]. As established previously for proteome analysis of purified mouse myelin [9,61], samples were re-run in a data acquisition mode without ion mobility separation of peptides (referred to as MS^{E}) to ensure the correct quantification of exceptionally abundant myelin proteins. Continuum LC-MS data were processed using Waters ProteinLynx Global Server (PLGS) and searched against a custom database compiled by adding the sequence information for *E. coli* chaperone protein ClpB and porcine trypsin to the UniProtKB/Swiss-Prot human proteome (release 2019-10, 20379 entries) and by appending the reversed sequence of each entry to enable the determination of false discovery rate (FDR). Precursor and fragment ion mass tolerances were automatically determined by PLGS and were typically below 5 ppm for precursor ions and below 10 ppm (root mean square) for fragment ions. Carbamidomethylation of cysteine was specified as fixed and oxidation of methionine as variable modification. One missed trypsin cleavage was allowed. Minimal ion matching requirements were two fragments per peptide, five fragments per protein and one peptide per protein. FDR for protein identification was set to 1% threshold.

For post-identification analysis including TOP3 quantification of proteins, ISOQuant [62]; software freely available at www.isoquant.net) was used as described previously [9,61]. Only proteins represented by at least two peptides (minimum length six amino acids, score ≥ 5.5, identified in at least two runs) were quantified as parts per million (ppm), i.e. the relative amount (w/w) of each protein in respect to the sum over all detected proteins. FDR for both peptides and proteins was set to 1% threshold and at least one unique peptide was required. Human myelin fractions and the corresponding white matter homogenates were assessed as five biological replicates (n = 5) each. The proteome analysis was repeated as an independent replicate experiment from the same protein fractions, resulting in ten LC-MS runs per condition. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [63] partner repository with dataset identifier PXD029727.

526

527 **Visualization of proteomic data**

528 Proteomic data were visualized and analyzed as in [9,61]. In more detail, heatmaps and scatter
529 plots were prepared in Microsoft Excel 2016 and GraphPad Prism 9. The area-proportional
530 Venn diagram was prepared using BioVenn [64].

531

532 **Gel electrophoresis and silver staining of gels**

533 Protein concentrations were determined using the DC Protein Assay kit (BioRad, Hercules,
534 CA, USA). Samples were diluted in 1 x SDS sample buffer with dithiothreitol and separated on
535 a 12% SDS-PAGE for 1 hr at 200 V using the BioRad system; gels were fixated overnight in
536 10% [v/v] acetic acid / 40% [v/v] ethanol, and then washed in 30% ethanol (2 x 20 min) and
537 ddH_2O (1 x 20 min). For sensitization, gels were incubated 1 min in 0.012% [v/v] $Na_2S_2O_3$ and
538 subsequently washed with ddH_2O (3 x 20 s). For silver staining, gels were impregnated for 20
539 min in 0.2% [w/v] $AgNO_3$/ 0.04% formaldehyde, washed with ddH_2O (3 x 20 s) and developed
540 in 3% [w/v] Na_2CO_3 / 0.02% [w/v] formaldehyde. The reaction was stopped by exchanging the
541 solution with 5% [v/v] acetic acid. Gels were kept in ddH_2O until documentation.

542

543 **Immunoblotting**

544 Immunoblotting was performed as described [37]. Primary antibodies were specific for connexin-
545 29 (GJC3, Invitrogen 34-4200, 1:500), tetraspanin-2 (TSPAN2, ProteinTech #20463-1-AP,
546 1:500), aspartoacylase (ASPA, ProteinTech #13244-1-AP, 1:500), tetraspanin-28 (CD81, BD
547 Biosciences-US #559517, 1:500), sirtuin 2 (SIRT2, Abcam #ab67299, 1:500), immunoglobulin
548 superfamily member 8 (IGSF8, Thermoscientific #PA5-71693, 1:500), carbonic anhydrase
549 (CA2, kind gift from Said Ghandour, 1:1000), myelin-associated oligodendrocyte basic protein
550 (MOBP, LS-Bio #LS-C164262/43727, 1:500), myelin basic protein (MBP, Serotec #PO2687,
551 1:500), cyclic nucleotide phosphodiesterase (CNP, Sigma #SAB1405637, 1:1000), proteolipid
552 protein (PLP/DM20, A431, kind gift from Martin Jung, 1:5000), myelin-associated glycoprotein
553 (MAG, clone 513, Chemicon #MAB1567, 1:500), myelin oligodendrocyte glycoprotein (MOG,
554 clone 8-18C5, kind gift from Christopher Linington 1:500), claudin-11 (CLDN11, Abcam
555 #ab53041, 1:500), protein-arginine deiminase 2 (PADI2, Proteintech #12110-1-AP, 1:1000),
556 tetraspanin-29 (CD9, Abcam #ab92726, 1:500), alpha-crystallin B chain (CRYAB, ProteinTech
557 #15808-1-AP, 1:500), peripheral myelin protein 2 (PMP2, ProteinTech #12717-1-AP, 1:500).
558 Appropriate secondary anti-mouse or anti-rabbit antibodies conjugated to HRP were from
559 Dianova (HRP goat-anti-mouse, #115-035-003, 1:5000; HRP goat-anti-rabbit, #111-035-003,
560 1:5000; HRP goat-anti-rat, #112-035-167, 1:5000). Immunoblots were developed using the
561 Enhanced Chemiluminescence Detection kit (Western Lightning Plus, Perkin Elmer, Waltham,
562 Massachusetts, USA) and the Super Signal™ West Femto Maximum Sensitivity Substrate

(Thermo Fisher Scientific, Rockford, Illinois, USA). Signal was detected using the Intas ChemoCam system (INTAS Science Imaging Instruments GmbH, Göttingen, Germany). Original immunoblots are provided in **Figure 2-source data 1**.

Immunohistochemistry

Paraffinized human optic nerves were cut in 5 µm sections using the microtome RM2155 (Leica, Wetzlar, Germany) and placed on 1 mm thick microscope slides (Marienfeld, #1000000, Lauda/Königshofen, Germany). Immunolabelling of the paraffinized cross sections was performed as follows: sections were incubated for 10 min at 60°C, deparaffinized in a series of incubations in xylol, xylol, xylol/isopropanol (1:1 ratio) for 10 min each, incubated in a series of steps in decreasing ethanol concentration (100%, 90%, 70%, 50%) for 5 min each, and finally washed in ddH_2O for 5 min. Afterwards, the sections were incubated for 5 min in 0.01 M citrate buffer (pH 6.0). Then, the sections with citrate buffer were microwaved at 600 W for 10 min. Finally, the slides were left to cool down, rinsed 1 x 5 min with 0.05 M Tris buffer (pH 7.6) containing 2% milk powder, and then blocked with 10% goat serum (Gibco/ThermoFisher Scientific #16210064, Waltham, Massachusetts, USA) diluted 1:4 in PBS (pH 7.4)/ 1% BSA. Primary antibodies were diluted in PBS/BSA and applied over night at 4°C. Samples were washed 3 x 5 min in Tris buffer with 2% milk powder (Frema Instant Magermilchpulver, granoVita, Radolfzell, Germany). Secondary antibodies were applied in incubation buffer (1:500 in PBS/BSA) with 4‘,6-diamidino-2-phenylindole (DAPI, ThermoFisher Scientific, Waltham, Massachusetts, USA, 1:2000). Slides were then rinsed 1 x 5 min with Tris buffer without milk powder and mounted using Aqua-Poly/Mount (Polysciences, Eppelheim, Germany). Antibodies were specific for peripheral myelin protein (PMP2; ProteinTech #12717-1-AP; 1:200) and human β-Tubulin 3 (TUJ1; BioLegends #MMS-435P; 1:500). Secondary antibodies were donkey-anti-mouse Alexa 555 (Invitrogen #A31570, 1:1000) and goat anti-rabbit Dylight 633 (Invitrogen #35562, 1:500). The labeled sections were imaged using the confocal microscope LSM880 (Zeiss, Oberkochen, Germany). The signal was collected with the objective Plan-Apochromat 40x/1.4 Oil DIC M27 using oil (Immersol™ 518 F, Zeiss, Oberkochen, Germany) and an additional zoom of 1.5. To observe the samples with the light source Colibri (Zeiss, Oberkochen, Germany), a FS90 filter was used. DAPI was excited at 405 and signal was collected between 431 nm – 495 nm. Alexa 555 was excited with a DPSS 561-10 laser at an excitation of 561 nm and signal was collected between 571 nm – 615 nm. Then, Dylight 633 was excited with a HeNe633 laser at an excitation of 633 nm and an emission between 647 nm – 687 nm. Finally, the MBS 488/561/633 beam splitter was used to detect Alexa 555 and Dylight 633 and MBS-405 for DAPI respectively. Images were processed with ImageJ software.

Retrieval of publicly available scRNA-seq datasets

Eight mouse and six human scRNA-seq datasets published between 2015 and early 2020 were collected for transcriptome analysis. Datasets were selected based on the number of cells designated as oligodendrocytes, and the reported health condition of specimen. Dataset expression matrices and, if available, corresponding metadata, were recovered for mouse datasets GSE60361 [32], GSE775330 [29], GSE113973 [16], GSE116470 [30], SRP135960 [31], GSE129788 [33], GSE130119 [28], and GSE140511 [18]. Human scRNA datasets were retrieved for the Single Cell Portal DroNC-Seq human archived brain [26], GSE97930 [25], GSE138852 [27], GSE118257 [17,28], GSE130119 [28], and syn21125841 [18]. For quality control, each of the retrieved datasets was analyzed using the Seurat R package (Version 3.1.4) [65]in an analysis pipeline including validating sequencing quality, filtering for outlier cells (as specified in **Figure 3-source data 1**), log-normalizing the expression matrix with a scale factor 10,000, high variable gene selection and data scaling, linear dimensionality reduction using principal component analysis (PCA), and neighboring embedding using uniform manifold approximation and projection (UMAP) to ensure accurate cell type annotation and to detect any potential batch effect. Marker genes used for annotating the oligodendrocyte lineage were *CSPG4, PCDH15*, *PDGFRa, PTPRZ1* and *VCAN* for OPCs, *BCAS1, ENPP6* and *GPR17* for NFO, and *CA2*, *CLDN11*, *CNP*, *CMTM5*, *MAG*, *MBP, MOBP*, *PLP1* and *SIRT2 for MOL.* Specific parameters applied to individual datasets and the number of recovered cells are listed in **Figure 3-source data 1**.

Merging and integration of scRNA-seq profiles of human and mouse MOL

Cells designated as MOL were subset from each dataset and focused for downstream analysis. Before merging human and mouse datasets, 16,255 mouse gene symbols were translated into human gene symbols using a reference gene list from Mouse Genome Informatics (The Jackson Laboratory; retrieved from www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt on 28.10.2020; The Jackson Laboratory; Version 6.16). 32,952 additional mouse gene symbols were translated into human gene symbols by capitalizing the lettering. Gene symbol synchronized human and mouse MOL profiles were first merged and proceeded with the general analysis pipeline for identifying possible batch effects. Principal component analysis was performed using the top 2,000 most variable genes, and UMAP analysis was performed with the top 20 principal components (PCs); the results implied that the different studies introduced the largest variability for data separation. For integrating all selected human and mouse datasets, the SCTransform[66] pipeline implemented in Seurat was applied. Each dataset underwent SCTransform normalization, and all datasets were integrated using 3,000 identified integration features. Principal component analysis was conducted downstream and UMAP calculation

637 was performed using the first 20 PCs. Cluster analysis was based on the k-nearest neighbors
638 (KNN) algorithm calculated with resolution 0.1, and clusters of differentially expressed genes
639 were calculated using the model-based analysis of single-cell transcriptomics (MAST)
640 algorithm (**Figure 4-source data 1**).
641
642 **Human and mouse transcriptome correlation analysis**
643 Transcriptome correlation analysis of human and mouse MOL scRNA-seq profiles was
644 performed using the van der Waerden (vdW) score-transformed average expression of
645 integration features (n = 3,000) in each dataset. Subsequently, human and mouse gene
646 average vdW scores were visualized using scatter plot (in **Figure 3-supplement 2**); Pearson
647 correlation was calculated for annotated known myelin genes and all genes, respectively.
648
649 **Gene ontology enrichment analysis**
650 The resulting cluster marker gene lists were input for gene ontology (GO) enrichment analysis
651 to detect potential regulated biological processes terms using the gprofiler2 R package
652 (Version 0.2.0)[67]. GO terms with an FDR-corrected $P < 0.05$ were considered as enriched and
653 visualized using EnrichmentMap and AutoAnnotate plugins in Cytoscape (Version 3.8.2)[68–70]
654
655 **Statistics and reproducibility**
656 Pie charts, heatmaps and scatter plots were prepared in Microsoft Excel 2016 and GraphPad
657 Prism 9. For the scatter plots, Pearson correlation and regression line were calculated via
658 GraphPad Prism 9. Relative sample proteomic profile distances were evaluated using
659 Pearson's correlation based on log2 transformed ppm values. scRNA-seq cluster marker
660 analysis was conducted using MAST algorithm. Data analysis and visualization was performed
661 using GraphPad Prism 9 and R software.
662
663 **Data availability statement**
664 The mass spectrometry proteomics data for human myelin are supplied as **Figure 1-source**
665 **data 1** and have been deposited to the ProteomeXchange Consortium via the PRIDE [63] partner
666 repository with dataset identifier PXD029727. The utilized publicly available mass
667 spectrometry proteomics data for mouse myelin are available at the ProteomeXchange
668 Consortium PRIDE partner repository with dataset identifier PXD020007. Identifiers of utilized
669 publicly available scRNA-seq datasets are given in **Figure 2-source data 1**.
670
671 **Code availability statement**
672 Code can be accessed at https://github.com/TSun-tech/Gargareta_etal
673
674

675 **References**
676
677 1. Nave, K.-A. & Werner, H. B. Myelination of the Nervous System: Mechanisms and
678 Functions. *Annual Review of Cell and Developmental Biology* **30**, 503–533 (2014).
679 2. Stadelmann, C., Timmler, S., Barrantes-Freer, A. & Simons, M. Myelin in the Central
680 Nervous System: Structure, Function, and Pathology. *Physiol Rev* **99**, 1381–1431 (2019).
681 3. Duncan, I. D., Hammang, J. P. & Trapp, B. D. Abnormal compact myelin in the myelin-
682 deficient rat: absence of proteolipid protein correlates with a defect in the intraperiod line.
683 *Proceedings of the National Academy of Sciences* **84**, 6287–6291 (1987).
684 4. Popko, B. *et al.* Myelin deficient mice: Expression of myelin basic protein and generation
685 of mice with varying levels of myelin. *Cell* **48**, 713–721 (1987).
686 5. Edgar, J. M. *et al.* Early ultrastructural defects of axons and axon–glia junctions in mice
687 lacking expression of Cnp1. *Glia* **57**, 1815–1824 (2009).
688 6. Snaidero, N. *et al.* Antagonistic Functions of MBP and CNP Establish Cytosolic Channels
689 in CNS Myelin. *CellReports* **18**, 314–323 (2017).
690 7. Morell, P., Lipkind, R. & Greenfield, S. Protein composition of myelin from brain and spinal
691 cord of several species. *Brain Research* **58**, 510–514 (1973).
692 8. Franz, T., Waehneldt, T. v, Neuhoff, V. & Wächtler, K. Central Nervous System Myelin
693 Proteins and Glycoproteins in Vertebrates: a Phylogenetic Study. *Brain Research* **226**,
694 245–258 (1981).
695 9. Jahn, O. *et al.* The CNS Myelin Proteome: Deep Profile and Persistence After Post-
696 mortem Delay. *Frontiers in Cellular Neuroscience* **14**, (2020).
697 10. Ackerman, S. D. & Monk, K. R. The scales and tales of myelination: using zebrafish and
698 mouse to study myelinating glia. *Brain Research* vol. 1641 79–91 (2016).
699 11. Ravi, V. & Venkatesh, B. The Divergent Genomes of Teleosts. *Annual Review of Animal*
700 *Bioscience* **6**, 47–68 (2017).
701 12. Yoshida, M. & Colman, D. R. Parallel Evolution and Coexpression of the Proteolipid
702 Proteins and Protein Zero in Vertebrate Myelin. *Neuron* **16**, 1115–1126 (1996).
703 13. Siems, S. B. *et al.* Proteome Profile of Myelin in the Zebrafish Brain. *Frontiers in Cell and*
704 *Developmental Biology* **9**, (2021).
705 14. Yang, S. *et al.* Patterns of insertions and their covariation with substitutions in the rat,
706 mouse, and human genomes. *Genome Research* **14**, 517–527 (2004).
707 15. Yeung, M. S. Y. *et al.* Dynamics of oligodendrocyte generation in multiple sclerosis. *Nature*
708 **566**, 538–542 (2019).
709 16. Falcão, A. M. *et al.* Disease-specific oligodendrocyte lineage cells arise in multiple
710 sclerosis. *Nature Medicine* **24**, 1837–1844 (2018).
711 17. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature*
712 **566**, 543–547 (2019).
713 18. Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-
714 dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nature*
715 *Medicine* **26**, 131–142 (2020).
716 19. Silva, J. C., Gorenstein, M. v., Li, G. Z., Vissers, J. P. C. & Geromanos, S. J. Absolute
717 quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Molecular and*
718 *Cellular Proteomics* **5**, 144–156 (2006).
719 20. Greenfield, S., Brostoff, S., Eylar, E. H. & Morell~, P. Protein Composition of Myelin of the
720 Peripheral Nervous System. *Journal of Neurochemistry* **20**, 1207–1216 (1973).
721 21. Uusitalo, M. *et al.* Human myelin protein P2: from crystallography to time-lapse membrane
722 imaging and neuropathy-associated variants. *The FEBS Journal* (2021)
723 doi:10.1111/febs.16079.
724 22. Campagnoni, A. T. Molecular Biology of Myelin Proteins from the Central Nervous System.
725 *The Journal of Neurochemistry* **51**, 1–14 (1988).
726 23. Waehneldt, T. v & Malotka, J. Comparative electrophoretic study of the wolfgram proteins
727 in myelin from several mammalia. *Brain Research* **189**, 582–587 (1980).
728 24. Ishii, A. *et al.* Human myelin proteome and comparative analysis with mouse myelin.
729 *PNAS* **106**, 14605–14610 (2009).
730 25. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in
731 the human adult brain. *Nature Biotechnology* **36**, 70–80 (2018).

26. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* **14**, 955–958 (2017).
27. Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience* **22**, 2087–2097 (2019).
28. Wheeler, M. A. *et al.* MAFG-driven astrocytes promote CNS inflammation. *Nature* **578**, 593–599 (2020).
29. Marques, S. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
30. Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030.e16 (2018).
31. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e22 (2018).
32. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
33. Ximerakis, M. *et al.* Single-cell transcriptomic profiling of the aging mouse brain. *Nature Neuroscience* **22**, 1696–1708 (2019).
34. Martins-De-Souza, D. Mass spectrometry-based proteomics to understand schizophrenia. *Brazilian Journal of Analytical Chemistry* **7**, 13–17 (2020).
35. Dhaunchak, A. S. *et al.* A proteome map of axoglial specializations isolated and purified from human central nervous system. *GLIA* **58**, 1949–1960 (2010).
36. Gopalakrishnan, G. *et al.* Lipidome and proteome map of myelin membranes. *Journal of Neuroscience Research* **91**, 321–334 (2013).
37. Patzig, J. *et al.* Septin/anillin filaments scaffold central nervous system myelin to accelerate nerve conduction. *eLife* **5**, (2016).
38. Soroor, F. *et al.* Revised subunit order of mammalian septin complexes explains their in vitro polymerization properties. *Molecular Biology of the Cell* **32**, 289–300 (2021).
39. Erwig, M. S. *et al.* Anillin facilitates septin assembly to prevent pathological outfoldings of central nervous system myelin. *eLife* **8**, (2019).
40. Birling, M.-C., Tait, S., Hardy, R. J. & Brophy, P. J. A Novel Rat Tetraspan Protein in Cells of the Oligodendrocyte Lineage. *The Journal of Neurochemistry* **73**, 2600–2608 (1999).
41. Terada, N. *et al.* The tetraspanin protein, CD9, is expressed by progenitor cells committed to oligodendrogenesis and is linked to β1 integrin, CD81, and Tspan-2. *GLIA* **40**, 350–359 (2002).
42. de Monasterio-Schrader, P. *et al.* Uncoupling of neuroinflammation from axonal degeneration in mice lacking the myelin protein tetraspanin-2. *GLIA* **61**, 1832–1847 (2013).
43. Shimokawa, N. & Miura, M. Rhombex-29, a Novel Gene of the PLP/ DM20-M6 Family Cloned From Rat Medulla Oblongata by Differential Display. *Journal of Neuroscience Research* **62**, 1–8 (2000).
44. Werner, H. B. *et al.* A critical role for the cholesterol-associated proteolipids PLP and M6B in myelination of the central nervous system. *GLIA* **61**, 567–586 (2013).
45. Nagy, J. I., Ionescu, A. v, Lynn, B. D. & Rash, J. E. Connexin29 and Connexin32 at Oligodendrocyte and Astrocyte Gap Junctions and in Myelin of the Mouse Central Nervous System. *J Comp Neurol* **464**, 356–370 (2003).
46. Kleopa, K. A., Orthmann, J. L., Enriquez, A., Paul, D. L. & Scherer, S. S. Unique distributions of the gap junction proteins connexin29, connexin32, and connexin47 in oligodendrocytes. *GLIA* **47**, 346–357 (2004).
47. Altevogt, B. M. & Paul, D. L. Four Classes of Intercellular Channels between Glial Cells in the CNS. *Journal of Neuroscience* **24**, 4313–4323 (2004).
48. Eiberger, J. *et al.* Expression pattern and functional characterization of connexin29 in transgenic mice. *GLIA* **53**, 601–611 (2006).
49. Abe, M. *et al.* PMP2/FABP8 induces PI(4,5)P2-dependent transbilayer reorganization of sphingomyelin in the plasma membrane. *Cell Reports* **37**, 109935 (2021).
50. Brostoff, S. W., Karkhanis, Y. D., Carlo, D. J., Reuter, W. & Eylar, E. H. Characterization of the Major Proteins of Rabbit Sciatic Nerve Myelin. *Brain Research* **86**, 449 (1975).
51. Kitamura, K., Suzuki, M., Suzuki, A. & Uyemura, K. The Complete Amino Acid Sequence of the P2 Protein in Bovine Peripheral Nerve Myelin. *FEBS Letters* **115**, (1980).

52. Dearmond, S. J., Deibler, G. E., Bacon, M., Kies, M. W. & Eng, L. F. The Journal of Histochemistry and Cytochemistry copyright II A Neurochemical and Immunocytochemical Study of P2 Protein in Human and Bovine Nervous Systems'. *The Journal of Histochemistry and Cytochemistry* **28**, 1275–1285 (1980).
53. Whitaker, J. N. The Protein Antigens of Peripheral Nerve Myelin. *Annals of Neurology* **9**, 56–64 (1981).
54. Zenker, J. *et al.* A role of peripheral myelin protein 2 in lipid homeostasis of myelinating Schwann cells. *GLIA* **62**, 1502–1512 (2014).
55. Motley, W. W. *et al.* De novo PMP2 mutations in families with type 1 Charcot-Marie-Tooth disease. *Brain* **139**, 1649–1656 (2016).
56. Hong, Y. bin *et al.* A Mutation in PMP2 Causes Dominant Demyelinating Charcot-Marie-Tooth Neuropathy. *PLoS Genetics* **12**, (2016).
57. Castelijns, B. *et al.* Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients. *Nature Communications* **11**, 301 (2020).
58. Kagawa, T., Mekada, E., Shishido, Y. & Ikenaka, K. Immune System-Related CD9 Is Expressed in Mouse Central Nervous System Myelin at a Very Late Stage of Myelination. *J. Neurosci. Res* **50**, 312–320 (1997).
59. Madhavarao, C. N. *et al.* Immunohistochemical Localization of Aspartoacylase in the Rat Central Nervous System. *Journal of Comparative Neurology* **472**, 318–329 (2004).
60. Erwig, M. S. *et al.* Myelin: Methods for purification and proteome analysis. in *Methods in Molecular Biology* vol. 1936 37–63 (Humana Press Inc., 2019).
61. Siems, S. B. *et al.* Proteome profile of peripheral myelin in healthy mice and in a neuropathy model. *eLife* **9**, (2020).
62. Distler, U., Kuharev, J. & Tenzer, S. Biomedical applications of ion mobility-enhanced data-independent acquisition-based label-free quantitative proteomics. *Expert Review of Proteomics* **11**, 675–684 (2014).
63. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research* **47**, D442–D450 (2019).
64. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - A web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, (2008).
65. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411–420 (2018).
66. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, (2019).
67. Raudvere, U. *et al.* G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191–W198 (2019).
68. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, (2010).
69. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research* **5**, (2016).
70. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504 (2003).

Acknowledgments

We thank S. Ghandour, M. Jung and C. Linington for antibodies, A. Fahrenholz, D. Hesse, R. Jung and B. Sadowski for technical help and the International Max Planck Research School for Genome Science (IMPRS-GS) for supporting V.-I.G. and S.B.S. We gratefully acknowledge all contributors of data and data platforms that facilitate access to data sets and promote open-science practices, including AD Knowledge portal, Gene Expression Omnibus (GEO), Mousebrain.org, Single-Cell Portal and the ProteomeXchange Consortium partner repository PRIDE.

Competing interests statement

The authors declare that no competing interests exist.

Funding

Deutsche Forschungsgemeinschaft (DFG) WE 2720/2-2 to HBW
Deutsche Forschungsgemeinschaft (DFG) WE 2720/4-1 to HBW
Deutsche Forschungsgemeinschaft (DFG) WE 2720/5-1 to HBW
European Research Council (ERC) Advanced Grant MyeliNano to KAN
The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Vasiliki-Ilya Gargareta	Investigation, Writing-review & editing
Josefine Reuschenbach	Investigation, Writing-review & editing
Sophie B. Siems	Investigation, Writing-review & editing
Ting Sun	Investigation, Supervision, Writing-review & editing
Lars Piepkorn	Investigation, Writing-review & editing
Carolina Mangana	Investigation, Writing-review & editing
Erik Späte	Investigation, Writing-review & editing
Sandra Goebbels	Investigation, Writing-review & editing
Inge Huitinga	Unpublished Resources, Writing-review & editing
Wiebke Möbius	Investigation, Writing-review & editing
Klaus-Armin Nave	Writing-review & editing
Olaf Jahn	Conceptualization, Investigation, Writing-review & editing
Hauke B. Werner	Conceptualization, Supervision, Writing-original draft, Writing-review & editing

Author ORCIDs

Vasiliki-Ilya Gargareta	http://orcid.org/0000-0002-6673-0714
Josefine Reuschenbach	http://orcid.org/0000-0002-8374-777X
Sophie B. Siems	http://orcid.org/0000-0002-7760-2507
Ting Sun	http://orcid.org/0000-0002-7104-7215
Wiebke Möbius	http://orcid.org/0000-0002-2902-7165
Klaus-Armin Nave	http://orcid.org/0000-0001-8724-9666
Olaf Jahn	http://orcid.org/0000-0002-3397-8924
Hauke B. Werner	http://orcid.org/0000-0002-7710-5738

Figure legends

Figure 1. Proteome analysis of human CNS myelin.

a Number and relative abundance of proteins identified in myelin purified from human normal-appearing white matter according to two data-independent acquisition (DIA) mass spectrometric modes (MS^E, $UDMS^E$). Note that $UDMS^E$ (blue) identifies a larger number of proteins in myelin but provides a lower dynamic range of quantification. MS^E (lime green) identifies fewer proteins, but the comparatively higher dynamic range of quantification provides information about the relative abundance of the exceptionally abundant myelin proteins proteolipid protein 1 (PLP1), myelin basic protein (MBP) and cyclic nucleotide phosphodiesterase (CNP). See **Figure 1-source data 1** for datasets. ppm, parts per million.

b Scatter plot comparing the log_2-transformed relative abundance of proteins identified in myelin by MS^E against their abundance as identified by $UDMS^E$. Data points highlighted in blue represent known myelin proteins, some of which are indicated. The correlation coefficient (r) was calculated for all proteins identified by MS^E (gray) and for known myelin proteins (blue). The regression line serves as navigational mean. Note that PLP, MBP, and CNP deviate the most from the regression line due to the limitations of $UDMS^E$ in the correct quantification of such exceptionally abundant myelin proteins.

c Heatmap comparing the relative abundance of known myelin proteins in purified myelin compared to white matter homogenate. Mass spectrometric quantification based on five biological replicates (M1, M2, M3, M4, M5) as the average of two technical replicates each. Each horizontal line displays the fold change (FC) of a known myelin protein of which the abundance is increased (magenta) or reduced (turqoise) in human myelin compared to its average abundance in white matter lysate plotted on a log_2 color scale. As to the technical quality of the proteomic data and the purity of the myelin fraction, also see Pearson's correlation coefficients in **Figure 1-supplement 2** and heatmap comparisons for marker proteins representing other cell types and organelles in **Figure 1-supplement 3**.

d Pie chart showing the relative abundance of proteins identified by MS^E in myelin purified from the human white matter. Relative abundance is given in percentage with relative standard deviation (% +/- RSD). Note that known myelin proteins constitute approximately 82% of the total myelin protein; proteins so far not known as myelin proteins constitute about 18%.

Figure 1-supplement 1. Electron micrograph of the myelin-enriched fraction.
Representative electron micrograph (EM) of the myelin-enriched light-weight membrane fraction purified from the normal appearing white matter of a human subject. Myelin sheaths are identified by their typical multimembrane structure; cellular compartments other than myelin were virtually undetectable.

Figure 1-supplement 2. Pearson's correlation for proteome analysis by MS^E and $UDMS^E$
Clustered heatmap of Pearson's correlation coefficients for protein abundance comparing myelin-enriched fraction and white matter homogenate by two data acquisition modes (MS^E in **a**; $UDMS^E$ in **b**). The sex of the human subjects is indicated above the heatmap (female, light gray; male, dark grey). Note that the groups 'lysate' (Control_L) and 'myelin' (Control_M) as expected form with a high correlation of > 0.9 and that the two technical replicates cluster together adjacently. See **Figure 1-source data 1** for datasets. n=5 biological replicates analyzed as two technical replicates each.

Figure 1-supplement 3. Heatmaps comparing the relative abundance of marker proteins in purified myelin versus white matter homogenate.
a-h Fold change (FC) comparing the abundance in myelin purified from human normal-appearing white matter with that in white matter homogenate according to the MS^E dataset of marker proteins selected on the basis of prior knowledge for neurons (**a**), synapses (**b**), astrocytes (**c**), microglia (**d**), antioxidative proteins (**e**), nuclear-encoded mitochondrial proteins (**f**), mitochondrial-encoded proteins (**g**), and constituents of the extracellular matrix (**h**). In the heatmap each horizontal line corresponds to the fold-change (FC) of the abundance of a protein in purified myelin compared to its average abundance in white matter homogenate plotted on a $\log_2$ color scale with increased (magenta) or decreased (turquoise) abundance in purified myelin. Heatmaps display five biological replicates (M1, M2, M3, M4, M5) as the average of two technical replicates each. See **Figure 1-source data 1** for datasets. Note that most marker proteins for cell types or compartments other than myelin are reduced in abundance or not detected (n.d., marked with a diagonally crossed field) in myelin.

Figure 2. Comparison of the protein composition of human and mouse CNS myelin.

a Silver-stained SDS-PAGE (0.9 µg protein load) of myelin purified from human normal-appearing white matter and C57Bl/6N-mouse brains. Note that the band patterns are roughly comparable but not identical. Gel shows n=3 biological replicates per species.

b Venn diagram comparing 81 selected known myelin proteins identified by MS^E and $UDMS^E$ in myelin purified from human white matter (blue) and C57Bl/6N-mouse brains (orange) as recently established using the same methods [9]. Note that most known myelin proteins were identified in myelin of both species, while multiple myelin proteins were identified in myelin of only one species.

c,d Scatter plots of the log_2-transformed relative abundance of proteins identified in human myelin by MS^E (**c**) or $UDMS^E$ (**d**) plotted against their relative abundance in mouse myelin as recently established using the same methods [9]. Correlation coefficients (r) were calculated for all proteins identified in human myelin (gray) or known myelin proteins (blue). Regression lines serve as navigational mean.

e Heatmap comparing the relative abundance of known myelin proteins identified by MS^E in human myelin with that in mouse myelin according to the same method [9]. Each horizontal line displays the fold-change (FC) of a protein in five biological replicates (M1- M5) of human myelin compared to its average abundance in CNS myelin of mice plotted on a log_2-color scale. Note that several proteins display higher abundance in human (blue) or mouse (orange) myelin, while others show approximately similar relative abundance (white).

f Immunoblot analysis confirms comparatively higher abundance in human myelin of PMP2, CRYAB, CD9 and PADI2, approximately equal abundance of PLP, CNP, SEPTIN2, SEPTIN7 and SEPTIN8, and comparatively higher abundance in mouse myelin of TSPAN2, GPM6B, GJC3, ASPA, MOBP, IGSF8, SIRT2, CLDN11, CA2, MAG and MOG, as implied by the MS^E analysis. Note that immunoblot-based comparison of the relative abundance of MBP across species is not straightforward because MBP displays one dominant isoform (18.5 kDa) in human CNS myelin but three main isoforms (14.0, 17.0 and 18.5 kDa) in mouse CNS myelin due to species-dependent alternative splicing. Blots show n=3 biological replicates per species. For immunohistochemistry detecting PMP2 in human optic nerve cross sections see **Figure 2-supplement 1**.

Figure 2-supplement 1. Detection of PMP2 in human CNS myelin by immunoblot and immunohistochemistry.

a Immunoblot analysis of myelin purified from human normal-appearing white matter (CNS), C57Bl/6N mouse brains (CNS) and C57Bl/6N mouse sciatic nerves (PNS) using antibodies specific for PMP2. Blot shows two biological replicates per condition. Note that PMP2 was readily detected in human CNS myelin and mouse PNS myelin but not in mouse CNS myelin. SIRT2 was detected as a control.

b Confocal micrograph of immunohistochemistry detecting PMP2 (green) and TUJ1 (magenta) in paraffin-embedded cross-sectioned optic nerves of a human subject. Note that PMP2 labels myelin sheaths (arrowheads), while TUJ1 labels neuron-specific beta-III tubulin, i.e. axonal microtubules in this cross section. Nuclear staining (DAPI) is in blue. Shown is one biological replicate representative of three biological replicates. Scale bar, 20 µm.

Figure 3. **Cross species scRNA-seq profile comparison of mature oligodendrocytes.**
a-c UMAP plot of the scRNA-seq profile of mature oligodendrocytes (MOL) integrated from previously established human **(b)** and mouse **(c)** datasets. In **(a)** cells contributed by distinct studies are highlighted in different colors; the corresponding references are given.
d,e UMAP feature plots highlighting expression of selected MOL marker genes **(d)** and transcription factors **(e)** in the integrated object comprising MOL of both humans (blue) and mice (orange).
f UMAP feature plots and violin plots exemplify genes that display preferential expression in MOL of humans (*PMP2*; *PADI2*) or mice (*GJC3*, *TSPAN2*, *CA2*).

Figure 3-supplement 1. Integrated scRNA-seq profiles of the oligodendrocyte lineage in humans and mice.

a,b UMAP plots of the scRNA-seq profiles of oligodendrocyte lineage cells integrated separately for humans (**a**) and mice (**b**) from previously published datasets. Cells contributed by distinct studies are highlighted in different colors; the corresponding references are given. **c-f** UMAP plots highlighting expression of marker genes for oligodendrocyte precursor cells (OPC; **c**), newly formed oligodendrocytes (NFO; **d**) and mature oligodendrocytes (MOL; **e**) in the integrated human (blue) and mouse (orange) datasets. Note that some genes display preferential expression in human (*PMP2*, *CD9*, *CRYAB*) or mouse (*GJC3*, *TSPAN2*, *CA2*) oligodendrocytes (**f**).

Figure 3-supplement 2. Cross species MOL transcriptome correlation.

Scatter plot comparing the van der Waerden (vdW) score-transformed average expression of 3000 integration features in human and mouse mature oligodendrocytes (MOL). Data points of 34 known myelin-related transcripts that encode proteins represented in the myelin proteome are highlighted in blue (22 of which annotated with gene name); 6 known myelin-related transcripts of which the protein product is not present in the myelin proteome are highlighted in orange; 4 myelination-related transcription factors are highlighted in green; other transcripts are displayed in gray. Pearson correlation coefficient is given for known myelin-related transcripts (as highlighted in blue) or all transcripts. Regression line represents fit of data for navigational purpose.

Figure 3-supplement 3. Comparisons of myelin proteome and MOL transcriptome

vdW scores calculated from the mean abundance of integration features (n = 3,000) in integrated scRNA-seq datasets of humans **(a,b)** and mice **(c,d),** and in the CNS myelin proteome in the corresponding species as acquired using the MS^E **(a,c)** and $UDMS^E$ **(b,d)** modes. vdW-normalized data were subjected to Pearson correlation analysis and visualized as scatter plot. Known myelin proteins are highlighted, and the distribution of normalized protein and RNA profiles are illustrated by density plots along the axes.

Figure 4. Human and mouse mature oligodendrocyte subpopulation analysis.

a UMAP plot showing five subpopulations of mature oligodendrocytes (MOL) identified upon integrating all human and mouse scRNA-seq datasets.

b Bubble plot showing the top 5 subpopulation marker genes. All cells in all clusters also express high levels of known myelin-related marker transcripts (*CNP, PLP1, MBP, MAG, CLDN11*).

c Relative proportion of mature oligodendrocyte subpopulations in humans and mice. Note that MOL of both species contribute to all subpopulations.

Figure 4-supplement 1. Gene ontology term topics enriched per subpopulation

Gene ontology (GO) terms of biological processes (small circles) were grouped as topics (large circles). Colors represent association with MOL clusters displayed in **Figure 4**. FDR < 0.05.

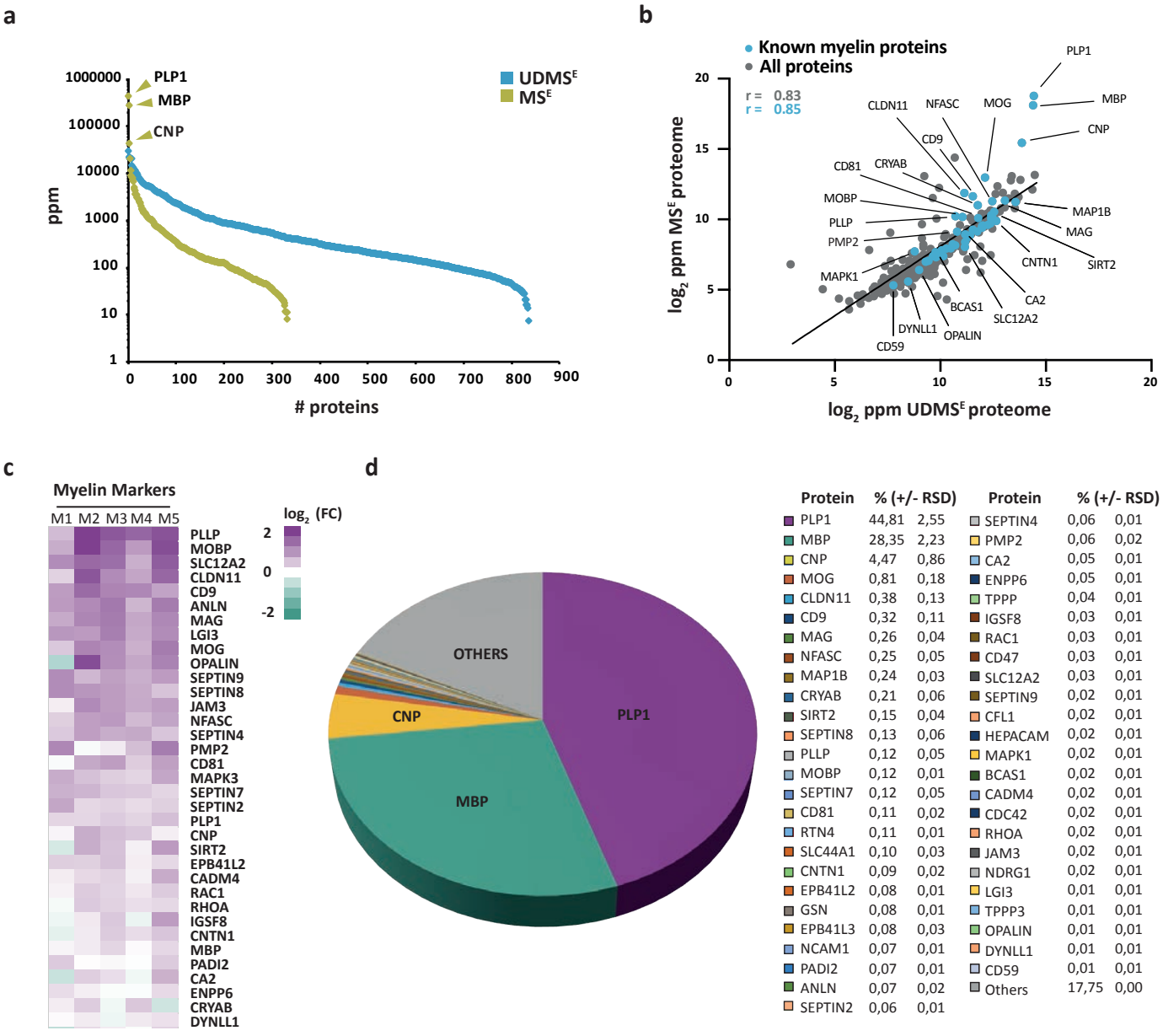
Figure 1-source data 1. Label-free quantification of proteins in human CNS myelin and white matter homogenate by two different data acquisition modes

Identification and quantification data of detected myelin-associated and homogenate proteins. Tryptic peptides derived from two technical replicates (replicate digestion) per five biological replicates were analyzed by LC-MS (10 runs per condition in total). Proteins (FDR < 1%; 2 peptides/protein) and peptides (FDR < 1%; ≥6 amino acids) were identified by database search against the UniprotKB/SwissProt mouse database using PLGS. Data were post-processed with the software package ISOQuant to calculate absolute in-sample amounts for each detected protein based on the TOP3 approach. Reported abundance values are defined as the relative amount of each protein in respect to the sum over all detected proteins (ppm: parts per million (w/w) of total protein). Typical contaminant proteins like albumin, hemoglobins, keratins, and trypsin were filtered. Tables are sorted by description (column D) in alphabetical order.

Figure 2-source data 1. Labeled original immunoblots

Figure 3-source data 1. Parameters applied for scRNA-seq individual dataset quality control and integrative analysis.

Figure 4-source data 1. MAST calculated marker genes from human and mouse integrated MOL subpopulations.



Protein	% (+/- RSD)		Protein	% (+/- RSD)	
PLP1	44,81	2,55	SEPTIN4	0,06	0,01
MBP	28,35	2,23	PMP2	0,06	0,02
CNP	4,47	0,86	CA2	0,05	0,01
MOG	0,81	0,18	ENPP6	0,05	0,01
CLDN11	0,38	0,13	TPPP	0,04	0,01
CD9	0,32	0,11	IGSF8	0,03	0,01
MAG	0,26	0,04	RAC1	0,03	0,01
NFASC	0,25	0,05	CD47	0,03	0,01
MAP1B	0,24	0,03	SLC12A2	0,03	0,01
CRYAB	0,21	0,06	SEPTIN9	0,02	0,01
SIRT2	0,15	0,04	CFL1	0,02	0,01
SEPTIN8	0,13	0,06	HEPACAM	0,02	0,01
PLLP	0,12	0,05	MAPK1	0,02	0,01
MOBP	0,12	0,01	BCAS1	0,02	0,01
SEPTIN7	0,12	0,05	CADM4	0,02	0,01
CD81	0,11	0,02	CDC42	0,02	0,01
RTN4	0,11	0,01	RHOA	0,02	0,01
SLC44A1	0,10	0,03	JAM3	0,02	0,01
CNTN1	0,09	0,02	NDRG1	0,02	0,01
EPB41L2	0,08	0,01	LGI3	0,01	0,01
GSN	0,08	0,01	TPPP3	0,01	0,01
EPB41L3	0,08	0,03	OPALIN	0,01	0,01
NCAM1	0,07	0,01	DYNLL1	0,01	0,01
PADI2	0,07	0,01	CD59	0,01	0,01
ANLN	0,07	0,02	Others	17,75	0,00
SEPTIN2	0,06	0,01			

Figure 1. Proteome analysis of human CNS myelin.

a Number and relative abundance of proteins identified in myelin purified from human normal-appearing white matter according to two data-independent acquisition (DIA) mass spectrometric modes (MS^E, $UDMS^E$). Note that $UDMS^E$ (blue) identifies a larger number of proteins in myelin but provides a lower dynamic range of quantification. MS^E (lime green) identifies fewer proteins, but the comparatively higher dynamic range of quantification provides information about the relative abundance of the exceptionally abundant myelin proteins proteolipid protein 1 (PLP1), myelin basic protein (MBP) and cyclic nucleotide phosphodiesterase (CNP). See **Fig.1-source data 1** for datasets. ppm, parts per million. **b** Scatter plot comparing the log_2-transformed relative abundance of proteins identified in myelin by MS^E against their abundance as identified by $UDMS^E$. Data points highlighted in blue represent known myelin proteins, some of which are indicated. The correlation coefficient (r) was calculated for all proteins identified by MS^E (gray) and for known myelin proteins (blue). The regression line serves as navigational mean. Note that PLP, MBP, and CNP deviate the most from the regression line due to the limitations of $UDMS^E$ in the correct quantification of such exceptionally abundant myelin proteins. **c** Heatmap comparing the relative abundance of known myelin proteins in purified myelin compared to white matter homogenate. Mass spectrometric quantification based on five biological replicates (M1, M2, M3, M4, M5) as the average of two technical replicates each. Each horizontal line displays the fold change (FC) of a known myelin protein of which the abundance is increased (magenta) or reduced (turqoise) in human myelin compared to its average abundance in white matter lysate plotted on a log_2 color scale. As to the technical quality of the proteomic data and the purity of the myelin fraction, also see Pearson's correlation coefficients in **Figure 1-Figure supplement 2** and heatmap comparisons for marker proteins representing other cell types and organelles in **Figure 1-Figure supplement 3**. **d** Pie chart showing the relative abundance of proteins identified by MS^E in myelin purified from the human white matter. Relative abundance is given in percentage with relative standard deviation (% +/- RSD). Note that known myelin proteins constitute approximately 82% of the total myelin protein; proteins so far not known as myelin proteins constitute about 18%.

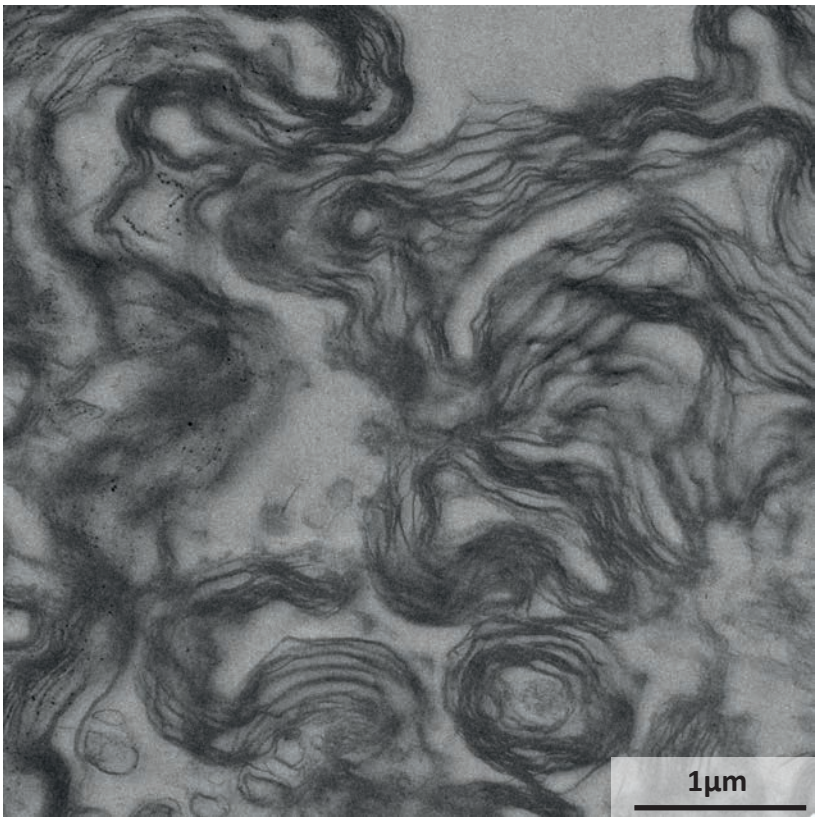


Figure 1 - figure supplement 1. Electron micrograph of the myelin-enriched fraction. Representative electron micrograph (EM) of the myelin-enriched light-weight membrane fraction purified from the normal appearing white matter of a human subject. Myelin sheaths are identified by their typical multimembrane structure; cellular compartments other than myelin were virtually undetectable.

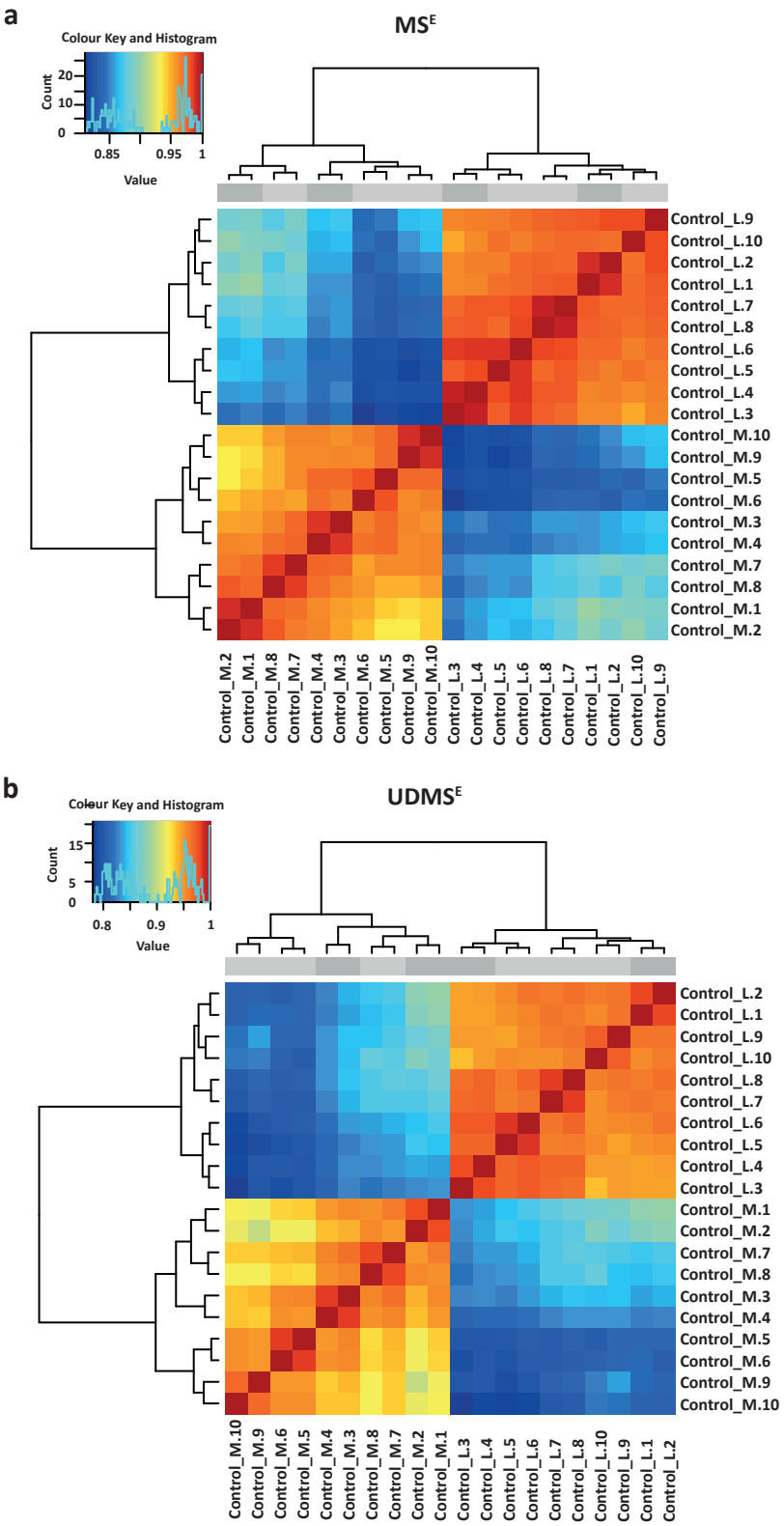


Figure 1 - figure supplement 2. Pearson's correlation for proteome analysis by MS^E and UDMS^E

Clustered heatmap of Pearson's correlation coefficients for protein abundance comparing myelin-enriched fraction and white matter homogenate by two data acquisition modes (MS^E in a; UDMS^E in b). The sex of the human subjects is indicated above the heatmap (female, light gray; male, dark grey). Note that the groups 'lysate' (Control_L) and 'myelin' (Control_M) as expected form with a high correlation of > 0.9 and that the two technical replicates cluster together adjacently. See **Fig1 - source data 1** for datasets. n=5 biological replicates analyzed as two technical replicates each.

Gargareta et al., Figure 1 - figure supplement 3

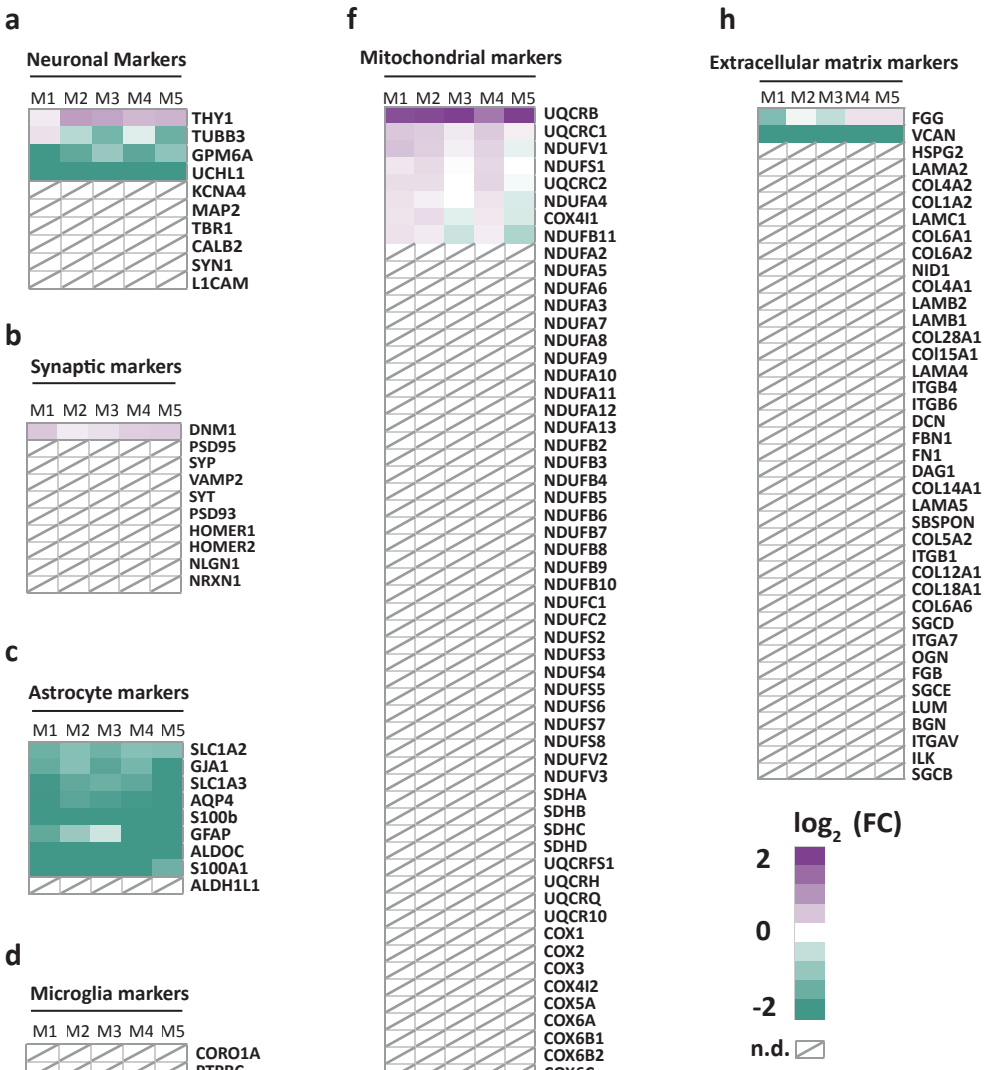


Figure 1 - figure supplement 3. Heatmaps comparing the relative abundance of marker proteins in purified myelin versus white matter homogenate.

a-h Fold change (FC) comparing the abundance in myelin purified from human normal-appearing white matter with that in white matter homogenate according to the MS[E] dataset of marker proteins selected on the basis of prior knowledge for neurons **(a)**, synapses **(b)**, astrocytes **(c)**, microglia **(d)**, antioxidative proteins **(e)**, nuclear-encoded mitochondrial proteins **(f)**, mitochondrial-encoded proteins **(g)**, and constituents of the extracellular matrix (h). In the heatmap each horizontal line corresponds to the fold-change (FC) of the abundance of a protein in purified myelin compared to its average abundance in white matter homogenate plotted on a $\log_2$ color scale with increased (magenta) or decreased (turquoise) abundance in purified myelin. Heatmaps display five biological replicates (M1, M2, M3, M4, M5) as the average of two technical replicates each. See **Fig1 - source data1** for datasets. Note that most marker proteins for cell types or compartments other than myelin are reduced in abundance or not detected (n.d., marked with a diagonally crossed field) in myelin.

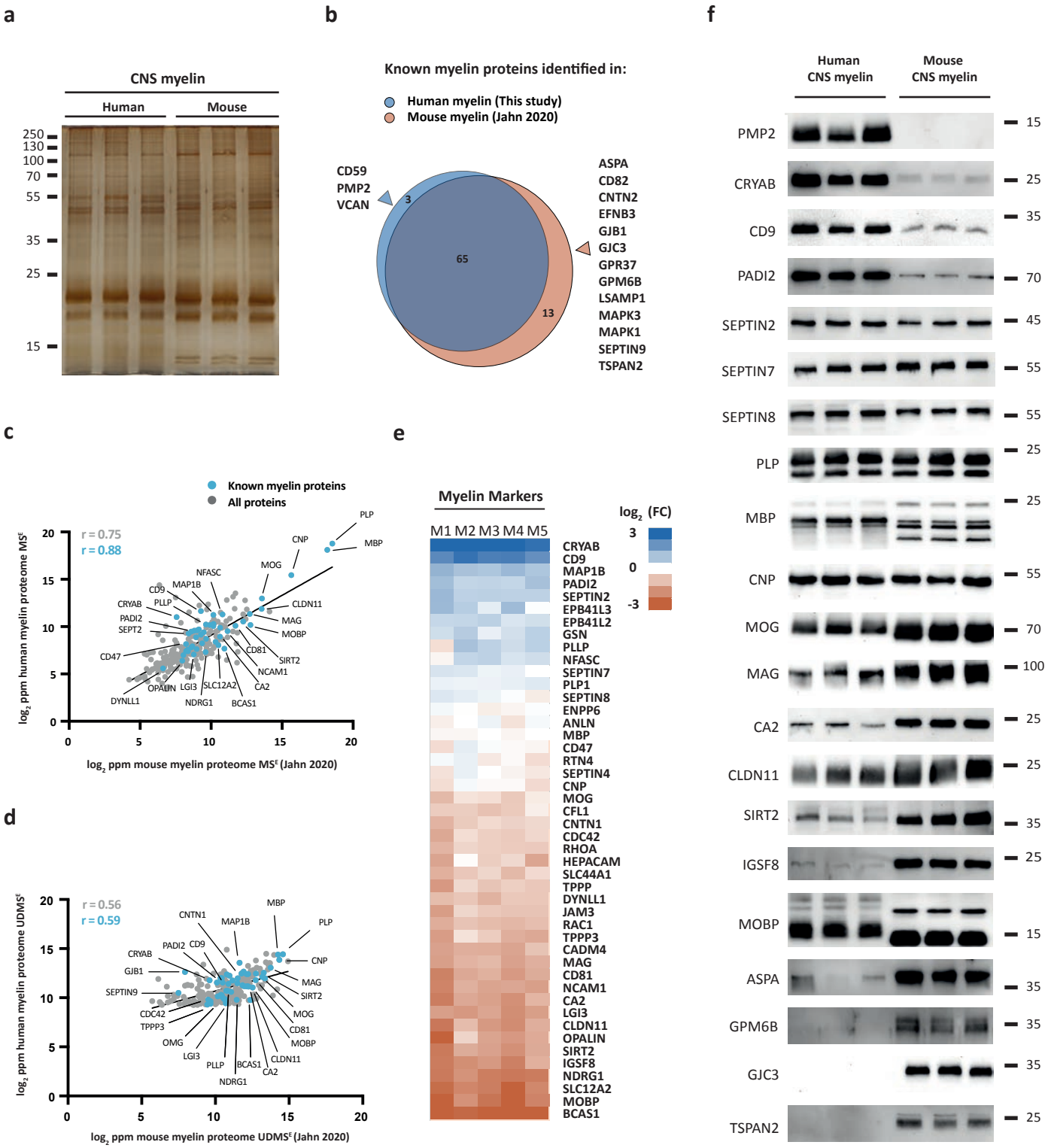


Figure 2. Comparison of the protein composition of human and mouse CNS myelin.
a Silver-stained SDS-PAGE (0.9 µg protein load) of myelin purified from human normal-appearing white matter and C57Bl/6N-mouse brains. Note that the band patterns are roughly comparable but not identical. Gel shows n=3 biological replicates per species. **b** Venn diagram comparing 81 selected known myelin proteins identified by MSE and UDMSE in myelin purified from human white matter (blue) and C57Bl/6N-mouse brains (orange) as recently established using the same methods 9. Note that most known myelin proteins were identified in myelin of both species, while multiple myelin proteins were identified in myelin of only one species. **c,d** Scatter plots of the log2-transformed relative abundance of proteins identified in human myelin by MSE (c) or UDMSE (d) plotted against their relative abundance in mouse myelin as recently established using the same methods. Correlation coefficients (r) were calculated for all proteins identified in human myelin (gray) or known myelin proteins (blue). Regression lines serve as navigational mean. e Heatmap comparing the relative abundance of known myelin proteins identified by MSE in human myelin with that in mouse myelin according to the same method 9. Each horizontal line displays the fold-change (FC) of a protein in five biological replicates (M1- M5) of human myelin compared to its average abundance in CNS myelin of mice plotted on a log2-color scale. Note that several proteins display higher abundance in human (blue) or mouse (orange) myelin, while others show approximately similar relative abundance (white). **f** Immunoblot analysis confirms comparatively higher abundance in human myelin of PMP2, CRYAB, CD9 and PADI2, approximately equal abundance of PLP, CNP, SEPTIN2, SEPTIN7 and SEPTIN8, and comparatively higher abundance in mouse myelin of TSPAN2, GPM6B, GJC3, ASPA, MOBP, IGSF8, SIRT2, CLDN11, CA2, MAG and MOG, as implied by the MSE analysis. Note that immunoblot-based comparison of the relative abundance of MBP across species is not straightforward because MBP displays one dominant isoform (18.5 kDa) in human CNS myelin but three main isoforms (14.0, 17.0 and 18.5 kDa) in mouse CNS myelin due to species-dependent alternative splicing. Blots show n=3 biological replicates per species. For immunohistochemistry detecting PMP2 in human optic nerve cross sections see Fig. 2 - figure supplement 1.

Gargareta et al., Figure 2 - figure supplement 1

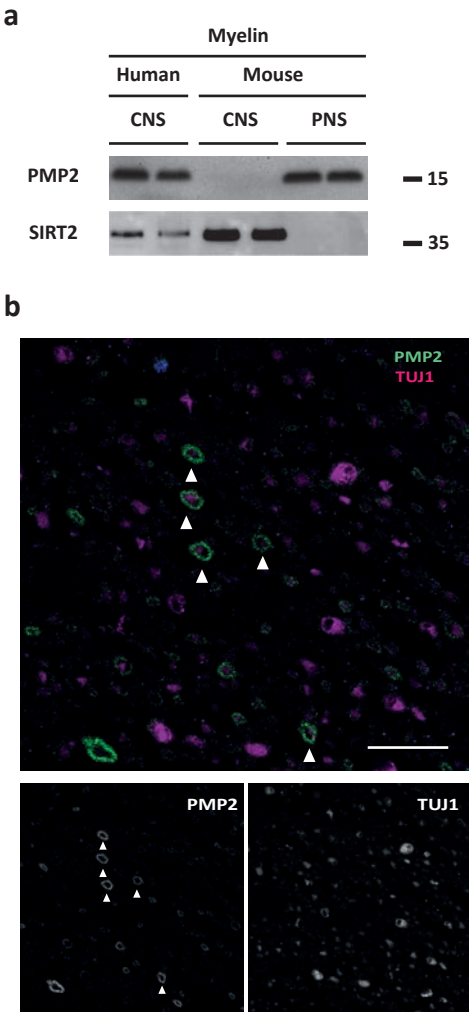


Figure 2 - figure supplement 1. Detection of PMP2 in human CNS myelin by immunoblot and immunohistochemistry.

a Immunoblot analysis of myelin purified from human normal-appearing white matter (CNS), C57Bl/6N mouse brains (CNS) and C57Bl/6N mouse sciatic nerves (PNS) using antibodies specific for PMP2. Blot shows two biological replicates per condition. Note that PMP2 was readily detected in human CNS myelin and mouse PNS myelin but not in mouse CNS myelin. SIRT2 was detected as a control. **b** Confocal micrograph of immunohistochemistry detecting PMP2 (green) and TUJ1 (magenta) in paraffin-embedded cross-sectioned optic nerves of a human subject. Note that PMP2 labels myelin sheaths (arrowheads), while TUJ1 labels neuron-specific beta-III tubulin, i.e. axonal microtubules in this cross section. Nuclear staining (DAPI) is in blue. Shown is one biological replicate representative of three biological replicates. Scale bar, 20 µm.

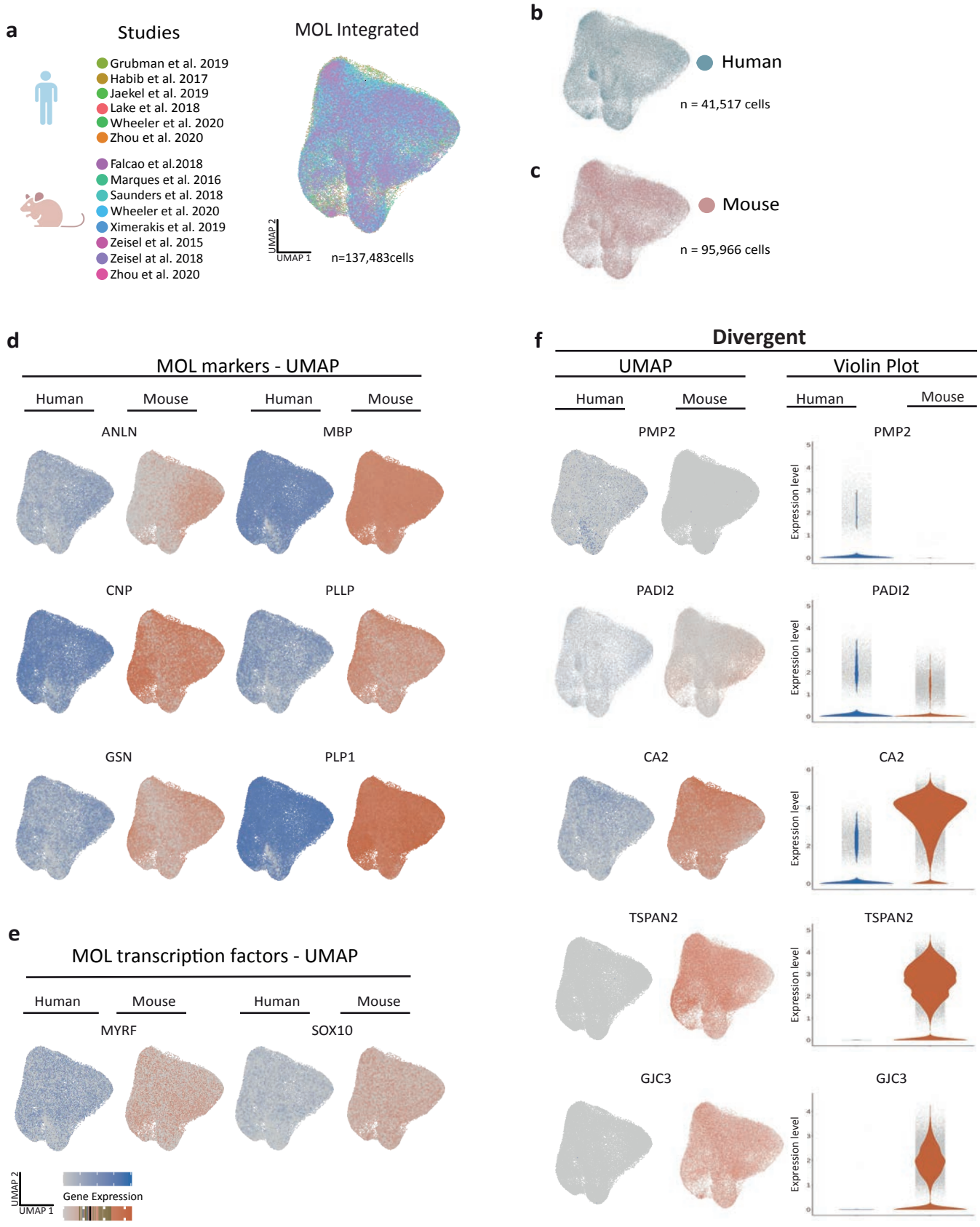


Figure 3. Cross species scRNA-seq profile comparison of mature oligodendrocytes.
a-c UMAP plot of the scRNA-seq profile of mature oligodendrocytes (MOL) integrated from previously established human (**b**) and mouse (**c**) datasets. In (**a**) cells contributed by distinct studies are highlighted in different colors; the corresponding references are given. **d,e** UMAP feature plots highlighting expression of selected MOL marker genes (**d**) and transcription factors (**e**) in the integrated object comprising MOL of both humans (blue) and mice (orange). **f** UMAP feature plots and violin plots exemplify genes that display preferential expression in MOL of humans (PMP2;PADI2) or mice (GJC3, TSPAN2, CA2).

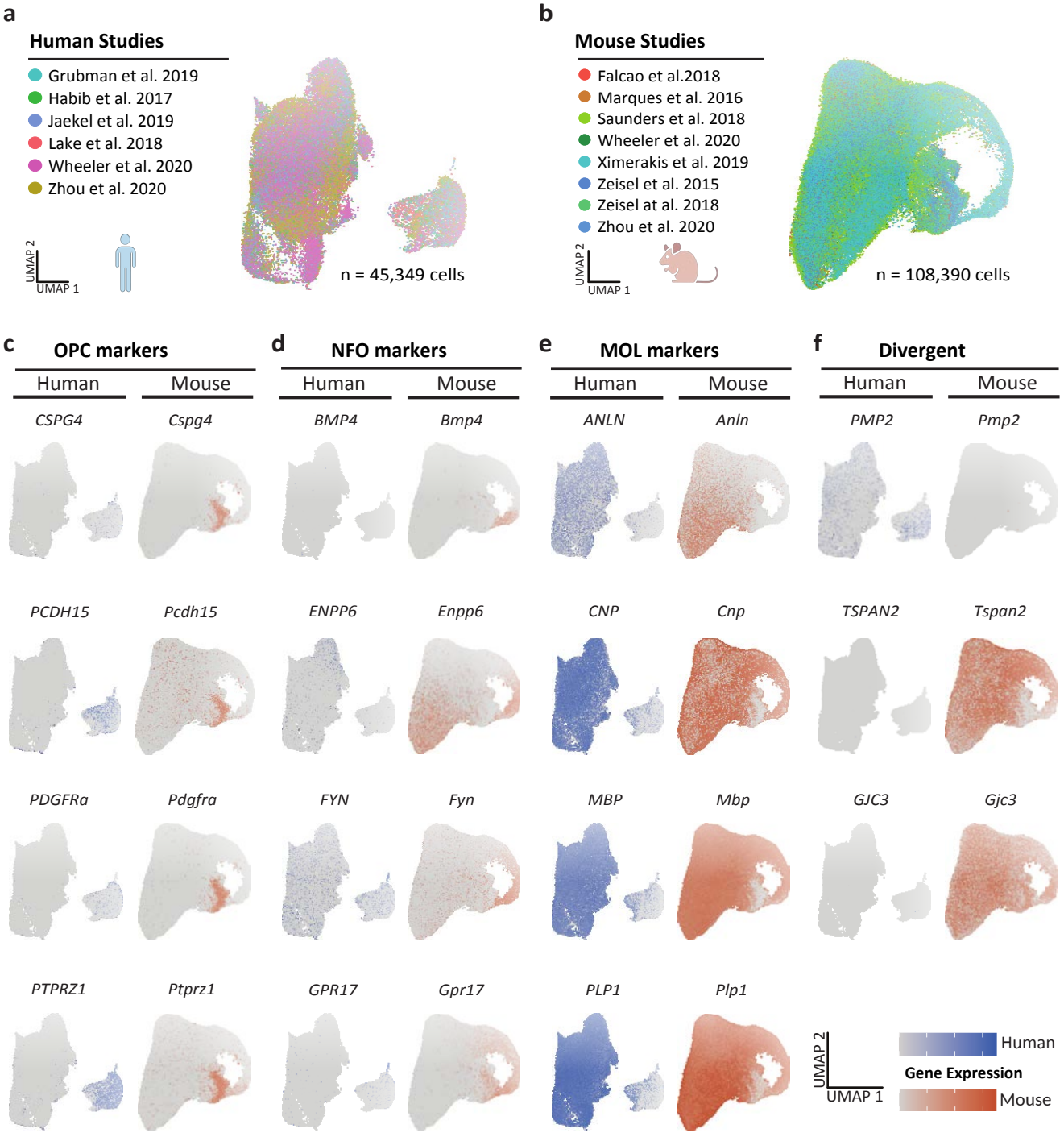


Figure 3 - figure supplement 1. Integrated scRNA-seq profiles of the oligodendrocyte lineage in humans and mice.
a,b UMAP plots of the scRNA-seq profiles of oligodendrocyte lineage cells integrated separately for humans (**a**) and mice (**b**) from previously published datasets. Cells contributed by distinct studies are highlighted in different colors; the corresponding references are given. **c-f** UMAP plots highlighting expression of marker genes for oligodendrocyte precursor cells (OPC; **c**), newly formed oligodendrocytes (NFO; **d**) and mature oligodendrocytes (MOL; **e**) in the integrated human (blue) and mouse (orange) datasets. Note that some genes display preferential expression in human (PMP2, CD9, CRYAB) or mouse (GJC3, TSPAN2, CA2) oligodendrocytes (**f**).

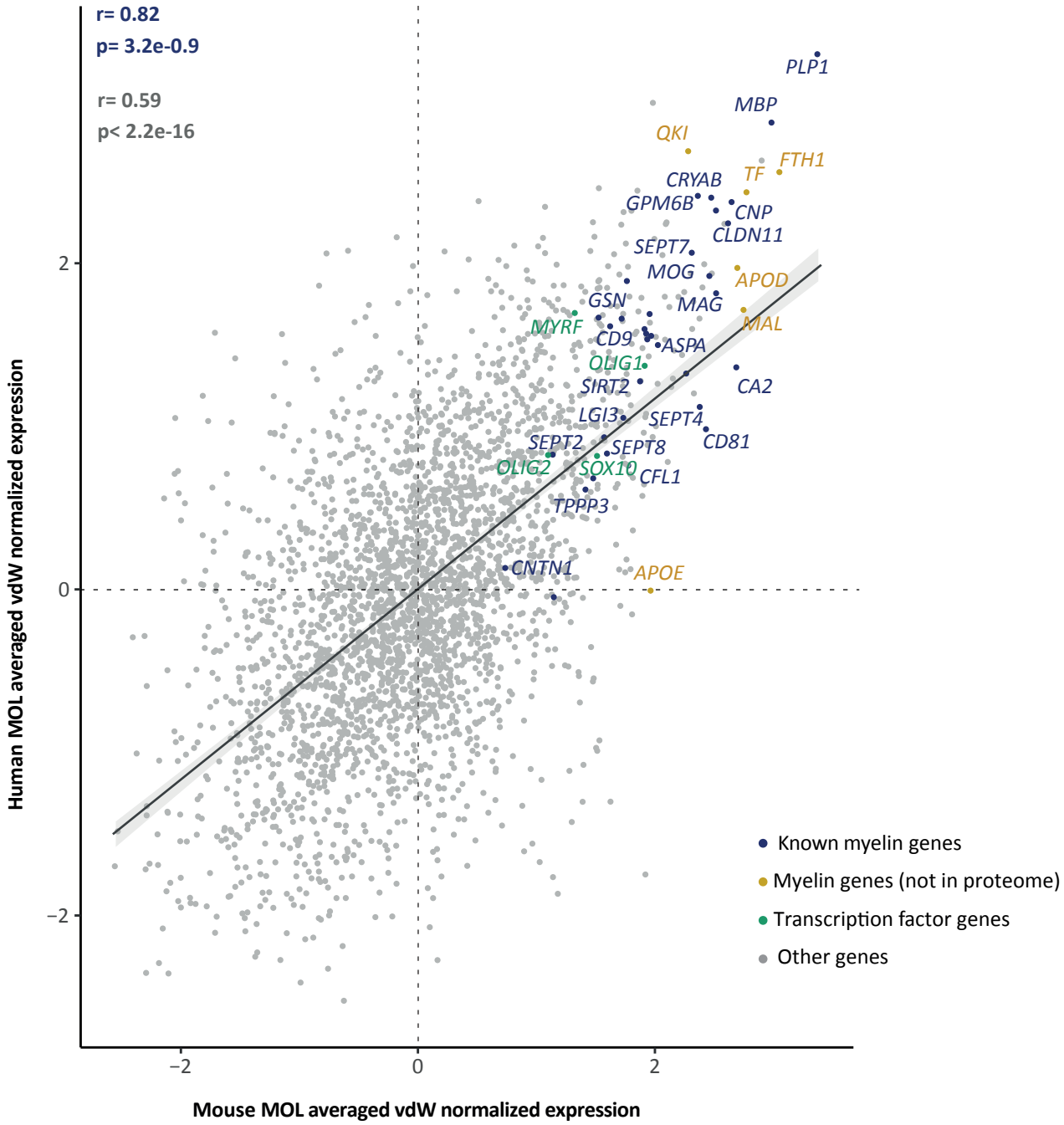


Figure 3 - figure supplement 2. Cross species MOL transcriptome correlation.
Scatter plot comparing the van der Waerden (vdW) score-transformed average expression of 3000 integration features in human and mouse mature oligodendrocytes (MOL). Data points of 34 known myelin-related transcripts that encode proteins represented in the myelin proteome are highlighted in blue (22 of which annotated with gene name); 6 known myelin-related transcripts of which the protein product is not present in the myelin proteome are highlighted in orange; 4 myelination-related transcription factors are highlighted in green; other transcripts are displayed in gray. Pearson correlation coefficient is given for known myelin-related transcripts (as highlighted in blue) or all transcripts. Regression line represents fit of data for navigational purpose.

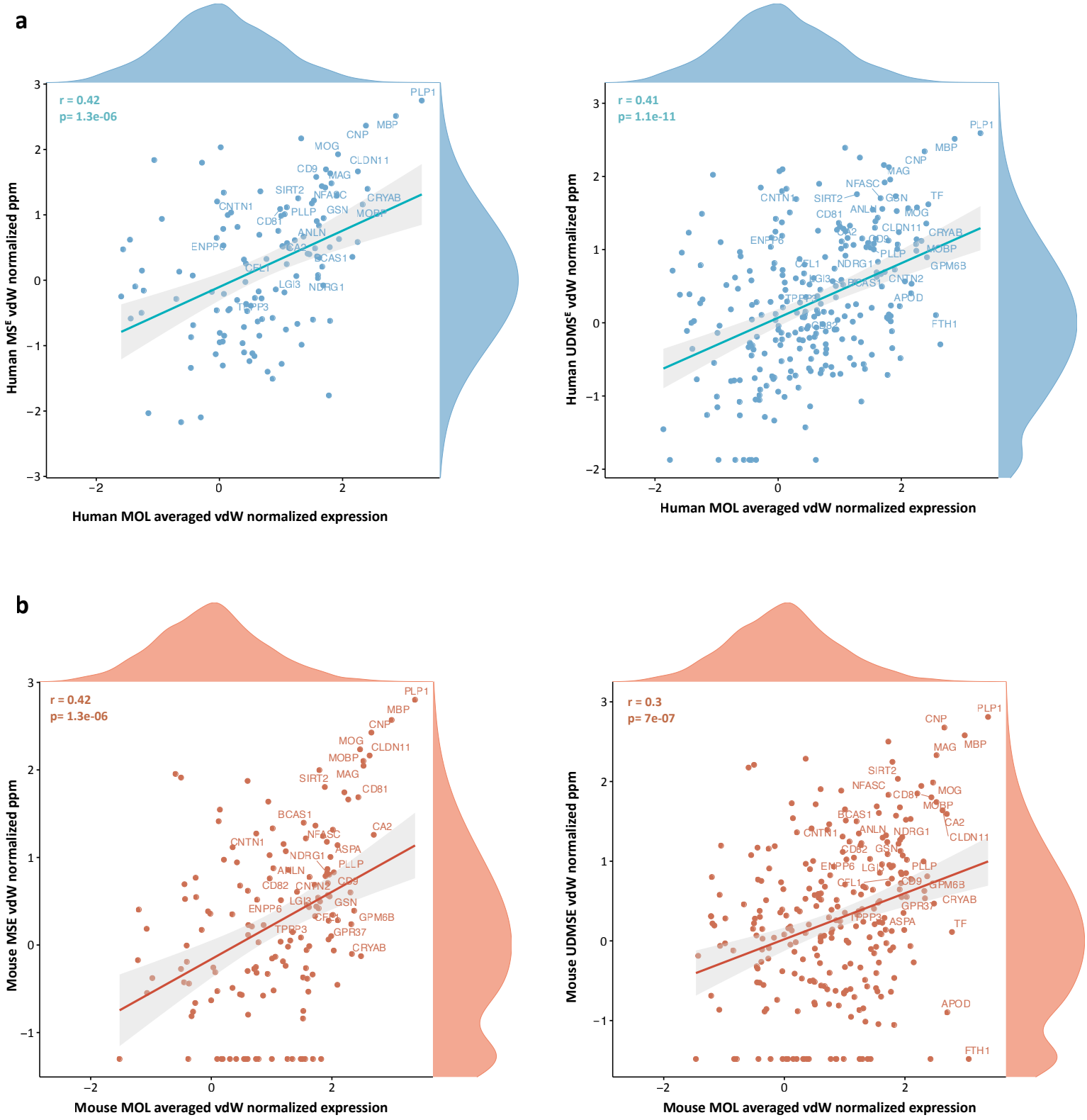


Figure 3 - figure supplement 2. Comparisons of myelin proteome and MOL transcriptome
vdW scores calculated from the mean abundance of integration features (n = 3,000) in integrated scRNA-seq datasets of humans (**a,b**) and mice (**c,d**), and in the CNS myelin proteome in the corresponding species as acquired using the MS^E (**a,c**) and UDMS^E (**b,d**) modes. vdW-normalized data were subjected to Pearson correlation analysis and visualized as scatter plot. Known myelin proteins are highlighted, and the distribution of normalized protein and RNA profiles are illustrated by density plots along the axes.

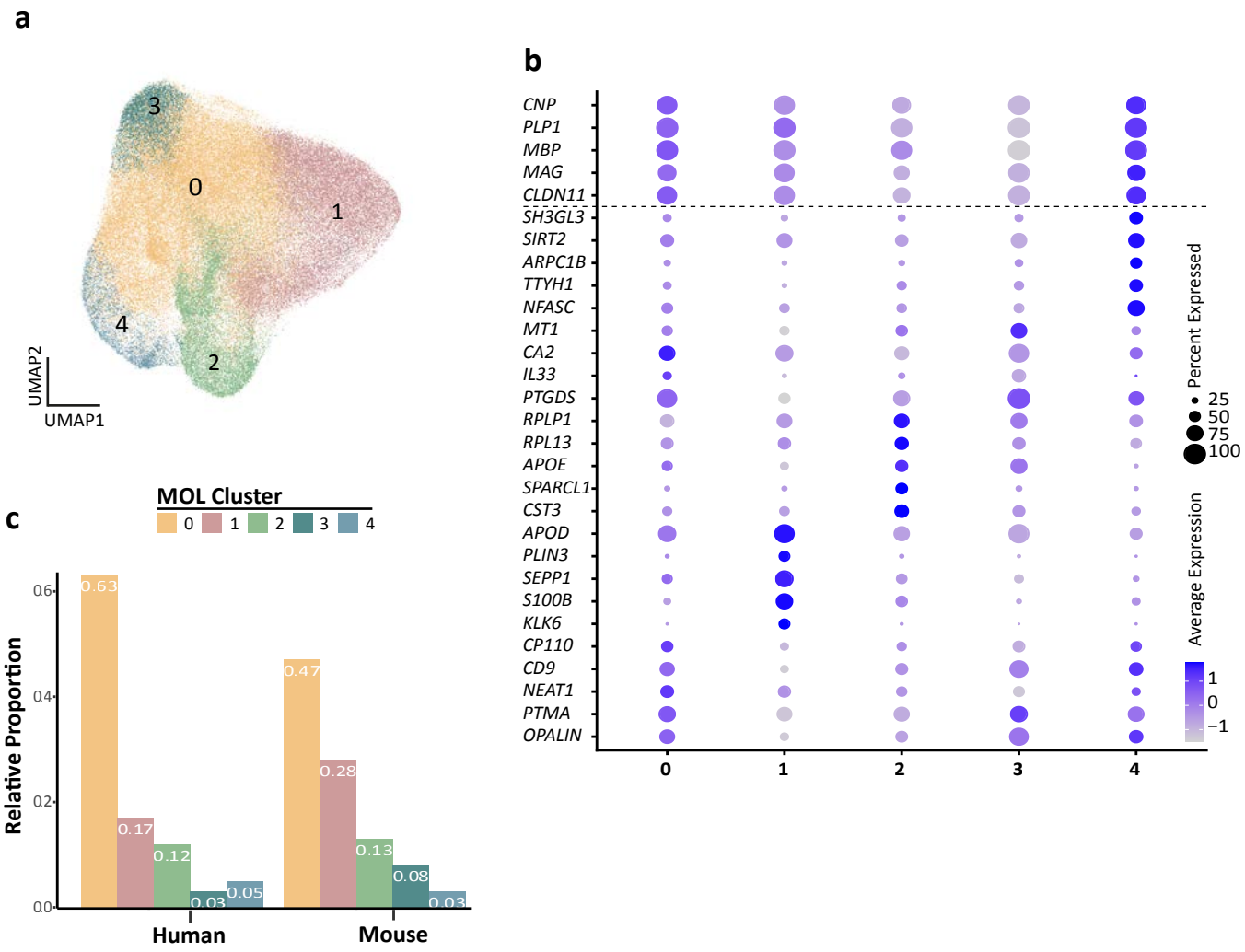


Figure 4. Human and mouse mature oligodendrocyte subpopulation analysis.
a UMAP plot showing five subpopulations of mature oligodendrocytes (MOL) identified upon integrating all human and mouse scRNA-seq datasets. **b** Bubble plot showing the top 5 subpopulation marker genes. All cells in all clusters also express high levels of known myelin-related marker transcripts (CNP, PLP1, MBP, MAG, CLDN11). **c** Relative proportion of mature oligodendrocyte subpopulations in humans and mice. Note that MOL of both species contribute to all subpopulations.

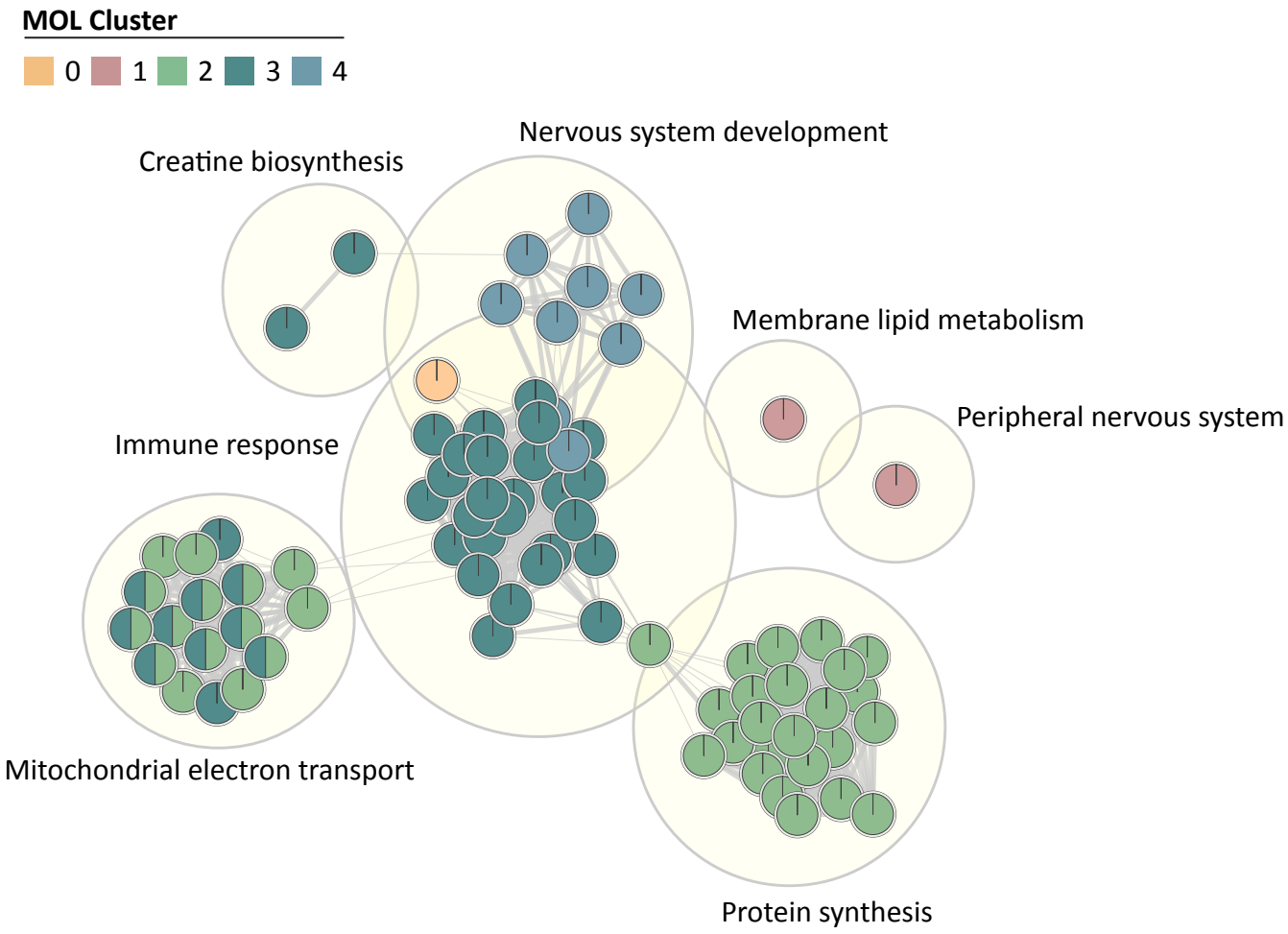


Figure 4 - figure supplement 1. Gene ontology term topics enriched per subpopulation
Gene ontology (GO) terms of biological processes (small circles) were grouped as topics (large circles). Colors represent association with MOL clusters displayed in **Figure 4**. FDR < 0.05.