



Supplementary Materials for

Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system

Sueli Marques^{1†}, Amit Zeisel^{1†}, Simone Codeluppi¹, David van Bruggen¹, Ana Mendenha Falcão¹, Lin Xiao³, Huiliang Li³, Martin Häring¹, Hannah Hochgerner¹, Roman A. Romanov^{1,2}, Daniel Gyllborg¹, Ana Muñoz Manchado¹, Gioele La Manno¹, Peter Lönnerberg¹, Elisa M. Floriddia¹, Fatemah Rezayee¹, Patrik Ernfors¹, Ernest Arenas¹, Jens Hjerling-Leffler¹, Tibor Harkany^{1,2}, William D. Richardson³, Sten Linnarsson^{1*}, Gonçalo Castelo-Branco^{1*}

¹ Laboratory of Molecular Neurobiology, Department Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

² Department of Molecular Neurosciences, Center for Brain Research, Medical University of Vienna, Austria

³ Wolfson Institute for Biomedical Research, University College London, Gower Street, London WC1E 6BT, United Kingdom

† equal contribution

*Correspondence to: sten.linnarsson@ki.se, goncalo.castelo-branco@ki.se

This PDF file includes:

Materials and Methods

Figs. S1 to S15

Additional Data Tables S1 to S3

Materials and Methods

Animals

Several mice strains were used for the single cell capture (Table S3). Briefly, wild-type (CD-1 or C57BL6) and transgenic mice between post-natal day 21 and 60 of both sexes were used (Fig S12). In the *Pdgfra*-H2BGFP knock-in mice (36), H2B-eGFP fusion gene is expressed under the promoter of the OPC marker, *Pdgfra*. The heterozygote mice express *Pdgfra* mainly in OPCs (37), but also in some extent in the early stages of OL differentiation, due to GFP half-life (Fig. 3). The *Lhx6* (Lim homeobox 6)-Cre transgenic mouse is a BAC transgenic mouse that express Cre under transcriptional control of *Lhx6* (38). To visualize the *Lhx6*⁺ cells this mouse line was crossed with the Cre-reporter line where tandem-dimer red fluorescent protein (tdRFP) is ubiquitously expressed under the ROSA26 locus (39). In the BAC transgenic mouse *5Htr3a*EGFP, EGFP is expressed under the control of the *Htr3a* promoter (GENSAT project, Rockefeller University, NY, USA). *Pdgfra*-Cre-RCE mice were used in P21-P30 and P60. *Pdgfra*-Cre-RCE mice with a C57BL/6 genetic background are a strain of mice expressing Cre recombinase under the control of a *Pdgfra* genomic DNA fragment (40) (The Jackson Laboratories, CA, USA). The mice was crossed with the reporter mice RCE:loxP-GFP (Gord Fishell, NYU Neuroscience Institute) to label the complete OL lineage (41). Complex motor wheel experiments were performed as described in (42). All experimental procedures performed followed the guidelines and recommendations of local animal protection legislation and were approved by the local committee for ethical experiments on laboratory animals (Stockholms Norra Djurförsöksnämnd in Sweden).

Tissue dissociation

Cells were isolated from ten distinct regions of the anterior-posterior and dorsal-ventral axis of the mouse juvenile and adult CNS, including grey matter (spinal cord/dorsal horn, substantia nigra and ventral tegmental area-/SN-VTA, amygdala, hypothalamic nuclei, zona incerta, hippocampus/dentate gyrus and CA1, and somatosensory cortex), white matter (corpus callosum) or both (striatum), and also adult CNS (somatosensory cortex, corpus callosum and dentate gyrus). Selected CNS regions were dissected in 300 μ m thick vibratome (VT1200 S, Leica) sections and cut out with scalpel. Tissue was then dissociated into a single cell suspension, as previously described in (6) with some modifications. Mice were perfused with oxygenated cutting solution (87 mM NaCl, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 26 mM NaHCO₃, 75 mM sucrose, 20 mM glucose, 1 mM CaCl₂, and 2 mM MgSO₄) and the brain was quickly dissected and kept in the same solution during sectioning on the vibratome. The selected regions were dissected from each slice, and dissociated using the Papain Neural dissociation kit from Miltenyi or the Worthington Papain dissociation system, following the manufacturer's instructions. Oxygenation and a short time of dissection were crucial to keep a high rate of survival in the cell suspension. The cell suspension was then filtered with 30 μ m filter (Partec) and myelin was removed using Myelin Removal beads (Miltenyi), the density gradient from the Worthington papain kit or using FACS to exclude debris. When required, cell suspension was also processed by FACS sorting for selection of GFP or RFP⁺ cells.

For the spinal cord, an extra gradient step was used after the strainer and final centrifugation step. The pellet was resuspended in 2ml Neurobasal-A medium including supplements and transferred onto an Optiprep Density Gradient (Sigma; cat#D1556), which consisted of 3 phases (Bottom phase: 160 μ l Optiprep + 340 μ l Neurobasal-A; middle phase: 220 μ l Optiprep + 1780 μ l Neurobasal-A; top phase: 75 μ l Optiprep + 925 μ l Neurobasal-A). The gradient was then centrifuged at 300g for 10min at 8°C and cells were located between the bottom and the middle phase.

FACS sorting

GFP or RFP+ cells were FACS sorted using a BD FACSaria III Cell Sorter B5/R3/V3 system. Cells were collected in cutting solution with 1%BSA and quickly prepared for capture on the C1 fluidigm system.

Cell capture and imaging

Cell suspension in a concentration of 600-1000 cells/ μ L was used. C1 Suspension Reagent was added (all 'C1' reagents were from Fluidigm, Inc.) in a ratio of 4 μ L to every 7 μ L cell suspension. 11 μ L of the cell suspension mix was loaded on a C1 Single-Cell AutoPrep IFC microfluidic chip designed for 10- to 17 μ m cells, and the chip was then processed on a Fluidigm C1 instrument using the 'mRNA Seq: Cell Load (1772x/1773x)' script (30 min at 4°C). The plate was then transferred to an automated microscope (Nikon TE2000E), and a bright-field image (20 \times magnification) was acquired for each capture site using μ Manager (<http://micro-manager.org/> (2)), which took <15 minutes. Quality of cells and control for doublets was performed after each experiment as described in (6).

Lysis, reverse transcription and PCR

C1 chips were processed as described in (6).

Tagmentation, Illumina high-throughput sequencing and molecule counts:

These steps were done as described in (6).

Clustering analysis

Cells identify as oligodendrocyte lineage from all tissues were collected into an oligodendrocyte dataset. Original full dataset include many experiments from different tissues and selection strategies. Most of the data was collected using unbiased sampling of cells, but some experiment included specific mice lines; 5HT3A-EGFP, LHX6-CRE-RFP for striatum and *Pdgfra*-H2B-GFP, *Pdgfra*-Cre-RCE for corpus callosum, hippocampus and cortex S1. The oligodendrocyte dataset is based on cells that identified as related to the lineage in the first round of clustering of the larger dataset (>20,000 cells). Being aware of significant RNA carryover and doublets contamination in the C1 Fluidigm system, we removed all cells that expressed (2 molecules or more) markers for

endothelial cells (Ly6c1, Cldn5), astrocytes (Agt, Htra1) or vascular smooth muscle (Acta2), which left us with 7487 cells. We use BackSpinV2 with the following parameters: splitlev = 5; NfUTURE1 = 500; NfUTURE = 50; N_to_backspin = 150; N_to_cells = 200; mean_tresh = 0.1; fdr_th = 0.3; min_gr_cells = 5; min_gr_genes = 3; which resulted in 16 clusters as shown in Figure S14. Manual inspection of those clusters results in merging three clusters and further exclusion of 2375 cells, which show either low quality or carryover from genes related to other types (marked in red bar in Fig S14). At the end of this step we had 5862 cells and 13 clusters. Another 790 cells were excluded after strong suspicion for mix between clusters, which was identified by using mature oligodendrocytes genes which were highly express in OPC, committed oligodendrocytes precursors or newly formed oligodendrocytes clusters. In practice we removed cells that have either one of the early markers *Bmp4*, *Itpr2*, *Cnksr3* and *Neu4* together with one of the late markers *Ctps*, *Mog* (>100) and *Klk6*. Also genes *Trf*, *Mal* (markers for mature oligodendrocytes) were used, cells that were classified to vascular and leptomeningeal cells, OPC, differentiation-committed oligodendrocytes precursors and newly formed oligodendrocyte population and express more than 80 molecules of those genes were excluded. Finally, cells that still express both astrocytes markers *Aqp4*, *Gja1* were excluded. The heatmap of the final 5072 cells in 13 clusters shown in Figure S1B.

Selection of cluster enriched genes and markers

When needed to select markers enriched genes, (for example as present in heatmap Figure S1) we use the following approach. For each gene *i* and cluster *j* calculate the ratios:

$$\mathbf{enrich}_{i,j} = \frac{1}{\langle k \in j \rangle} \sum_{k \in j} E_{i,k} / \frac{1}{N} \sum_k E_{i,k}; \quad \mathbf{posfrac}_{i,j} = \frac{1}{\langle k \in j \rangle} \sum_{k \in j} I(E_{i,k} > 0)$$

where $E_{(i,k)}$ is the expression of gene *i* in cell *k*. Those quantities represent the molecule enrichment in the cluster and the fraction of cells express the gene in that cluster respectively. We then combine the two with varies the weight given to the fraction of positive cells. Rank the gene for every cluster by the score:

$$[S_{(i,j)} = \text{enrich}_{(i,j)} \times [posfrac]_{(i,j)}]^{\text{power}}$$

Where “power” sets the weight for the fraction of positive cell in the cluster. We used power=0,0.5,1 to rank the genes in every cluster and then use the top X genes as most enriched. Specifically the dendrogram in Figure 1 was build based on 100 top genes per cluster as describe below. The gene selection for tSNE analysis (Figure2) we use the top 30 best markers and remove genes that do not express in 20% of in any of the populations.

Dendrogram construction and split point list

To build dendrogram in Figure 1C we first selected genes that are enriched for the cluster so that each cluster can be represented in a balanced way. This was done as described in

previous section, ranking the top enriched genes and selecting the top 100 from every power parameter. This resulted in 923 genes, which used to calculate the correlation (after log2+1 transform) between clusters. The correlation used as distance for average linkage method procedure to cluster and construct a tree. To better understand the tree shape we looked if there are specific genes that explain each of the tree junctions. To do this we use two different simple criteria the difference in average and the difference in fraction of positive (>0) cells. Each junction defines left and right groups and we calculated the above score at the single-cell level. We show the top 50 markers for every junction in Supporting File Supplementary Excel S1.

tSNE projections

We use tSNE (43) to visualize the dataset in 2D. tSNE projections shown in Figure 2A were calculated on 5072 cells as data points after selecting 334 clustered enriched genes as described above. We use 40 principle components, perplexity=80, epsilon (initial learning rate)=100, number of iterations 1000 and correlation as distance measure.

Pseudo-time analysis

Since we observed a smooth transition from OPC to myelin-forming oligodendrocytes we decided to look at the genes that are changing along this maturation process. Since the directionality along the main cloud of data points is well defined we manually draw a path from committed oligodendrocytes precursors to mature oligodendrocytes. Next, each point (closer than 10 units) was projected on the line assigning it a pseudo-time number. This creates an order of the data point and allows looking at the profile of every gene as a function of the pseudo time parameter, which reflects the stage in the maturation process. Pseudo-time dependent profiles were smoothed using running average and window of 100 points. To score the profiles we first transform each profile to a fold change minus one (FC-1) relative to the first 10 data points ($t=0$). This is done to allow a fair comparison of all genes. We score the profiles based on the integral (area between the line and the zero) and the number of times the line cross the zero (how fluctuating is the line). Profiles with the top 30th percentile area below the curve, which also cut the zero less than 10 times, are shown.

Unsupervised ordering of cells by Monocle

The single cell data was processed using Monocle (v1.2.0), as described in (44). Single cell data was ordered by Monocle according to developmental progress using an unsupervised clustering algorithm wherein the cells are represented as a single point in a n-dimensional space ($n=$ gene number), which is reduced to two dimensions using Independent Component Analysis (ICA). Next, a minimum spanning tree is constructed on the two-dimensional reduced space. The ordering algorithm then places cells on the diameter path in the order plotted in the reduced space. Cells that do not fall on the path are ordered in a way that minimizes expression differences with neighboring cells. Unsupervised clustering using 383 genes obtained from the previously performed

BackSpinV2 algorithm resulted in an unsupervised ordering of cells in accordance with expression state.

Immunohistochemistry

Wild-type mice at post-natal day 7, 21 and 90 or *Pdgfra*-H2B-GFP mice with 21 days were perfused with PBS followed by PFA 4% and brain were dissected and transferred to cold PFA 4%, where they were incubated for 2h, at 4°C. The brains were then transferred and kept overnight in a 30% sucrose solution. The tissues were embedded into OCT (Tissue-Tek) and sliced with 14µm thickness.

For the complex wheel experiments (25), free-floating sections from 4 non-runners and 6 runners after 2 days in the complex wheel were used. Each replicate had 3 free floating coronal sections, all sections within the same rostral-caudal position. Mice were perfused with 4% PFA in DEPC treated PBS, post fixed overnight and then sectioned 25µm in thickness, and stored in 50% glycerine PBS (DEPC treated).

Slides/sections were quickly boiled in antigen retrieval (Dako, S1699) and stand in the antigen solution until cooling down. They were then permeabilized in 1×PBS/0.3% Triton (Millipore) 3× 5 minutes and blocked for 1 hour in 1×PBS/0.3% Triton/5% donkey serum (Sigma, D9663) at room temperature and incubated overnight at 4°C with primary antibodies (GFP (Abcam, ab13970, chicken 1:2000), Sox10 (Santa cruz, cs-17342, goat 1:100), Tcf7l2 (Millipore, 05-511, mouse 1:200), Itpr2 (Millipore, AB3000, rabbit 1:40), DyLight 488 Labeled Lycopersicon Esculentum (Tomato) Lectin (Vector Labs, DL-1174, 1:100), Col1a1 (Abcam, ab21286, 1:50); Grm3 (Abcam, ab140741, Rabbit), CC1 (anti-APC; Millipore, OP80, Mouse 1:100) and Pdgfra (R&D, AF1062, Goat 1:200)) diluted in blocking solution. After washing the slides/sections 3x 5min with PBS1x, secondary antibodies (Invitrogen, 488 green 1:500, 555 red 1:1000 and 647 white 1:250) diluted in blocking solution were added and incubated for 1 hour at room temperature. Thereafter, slides were mounted with mounting medium containing DAPI (Vector, H-1200) and kept at 4°C until further microscopic analysis.

Microscopy and analysis

Combined images of DAPI, Alexa 555, GFP or Alexa 488 and Alexa 647, spanning the somatosensory cortex, CA1 and corpus callosum were obtained in a Zeiss LSM700 Confocal. For P7-P90 quantifications, 3 animals were used in each timepoint and 4-5 slices were photographed per animal. A total of 4 photos for CA1 and 6 for the corpus callosum and SsCx were counted. All the countings were normalized to the area analysed in each photo. % of Tcf7l2+ or Itpr2+ cells out of GFP or Sox10 were then calculated. For the complex wheel paradigm, 3 free floating coronal sections within the same rostral-caudal position were counted per animal. % Sox10+Itpr2+ cells in the corpus callosum of 4 non-runners and 6 runners were quantified.

Single-molecule fluorescence RNA in situ hybridization (smFISH)

smFISH was carried out as previously described (6) with minor modifications. Postnatal day 21 wild type CD1 were perfused with 4% cold PFA. The brains were then collected, embedded in Tissue-Tek OCT (Sakura), frozen on dry ice and stored at -80°C until used. The tissue sections were permeabilization for 10 min with Methanol at -20°C and

hybridisation for 4hrs with 250 nM fluorescent label probes (Biosearchtech, Petaluma, CA, USA) at 37°C and counterstained with DAPI (Thermo). The sections were mounted with pro-long gold (Thermo) and image stacks (0.3 μ m distance) were acquired using a customized automated scanning microscope controlled by μ Manager. Images were analyzed using a custom python script. After background removal a Laplacian-of-Gaussian was used to enhance the RNA dots that were then identified by computing the extended maxima transform of the flattened image.

GO analysis

For the complete gene ontology, we rank all genes by the average expression in population A vs B (related to dendrogram obtained by linkage clustering). The lists of 50 top genes in each branch were then applied to DAVID v6.7, the Database for Annotation, Visualization and Integrated Discovery (46, 47). This tool is a theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, PIR and Uniprot/SwissProt. Functional categories were ranked based on co-occurrence with sets of genes in a gene list, which unravels new biological processes associated with cellular functions and pathways. Any given gene is associating with a set of annotation terms. If genes share similar set of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics (46, 47).

36. R. A. Klinghoffer, T. G. Hamilton, R. Hoch, P. Soriano, An allelic series at the PDGFalphaR locus indicates unequal contributions of distinct signaling pathways during development. *Dev Cell* **2**, 103-113 (2002).
37. L. E. Clarke *et al.*, Properties and fate of oligodendrocyte progenitor cells in the corpus callosum, motor cortex, and piriform cortex of the mouse. *J Neurosci* **32**, 8173-8185 (2012).
38. M. Fogarty *et al.*, Spatial genetic patterning of the embryonic neuroepithelium generates GABAergic interneuron diversity in the adult cortex. *J Neurosci* **27**, 10935-10946 (2007).
39. H. Luche, O. Weber, T. Nageswara Rao, C. Blum, H. J. Fehling, Faithful activation of an extra-bright red fluorescent protein in "knock-in" Cre-reporter mice ideally suited for lineage tracing studies. *Eur J Immunol* **37**, 43-53 (2007).
40. K. Roesch *et al.*, The transcriptome of retinal Muller glial cells. *J Comp Neurol* **509**, 225-238 (2008).
41. J. H. Seo *et al.*, Oligodendrocyte precursor cells support blood-brain barrier integrity via TGF-beta signaling. *PLoS one* **9**, e103174 (2014).
42. I. A. McKenzie *et al.*, Motor skill learning requires active central myelination. *Science* **346**, 318-322 (2014).
43. L. J. P. v. d. M. a. G. E. Hinton, Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* **9(Nov)**, 2579-2605 (2008).
44. C. Trapnell *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32**, 381-386 (2014).

45. A. Zeisel *et al.*, Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142 (2015).
46. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
47. W. Huang da, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
48. Y. Zhang *et al.*, An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-11947 (2014).

Marques and Zeisel et al, Figure S1

A

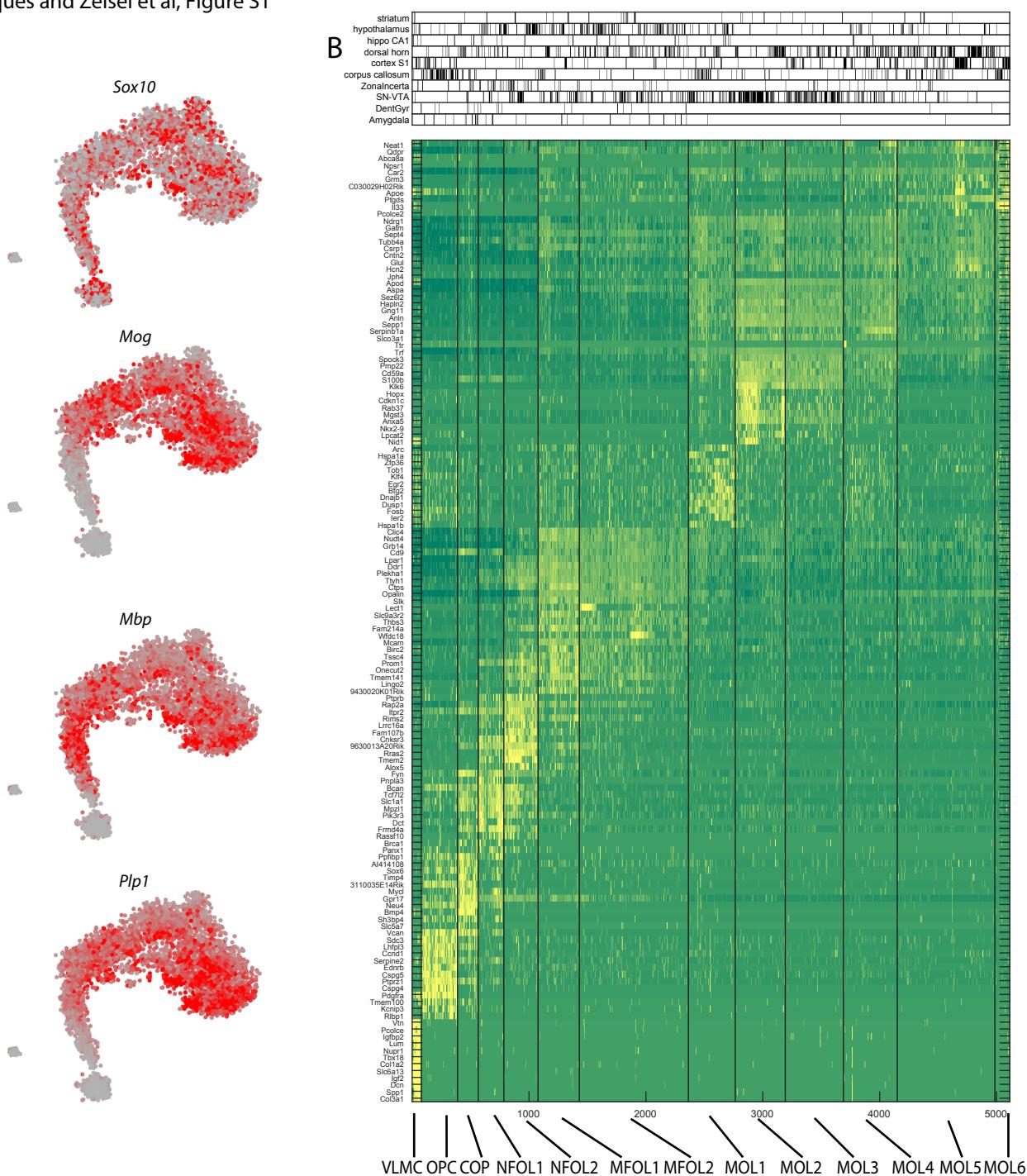
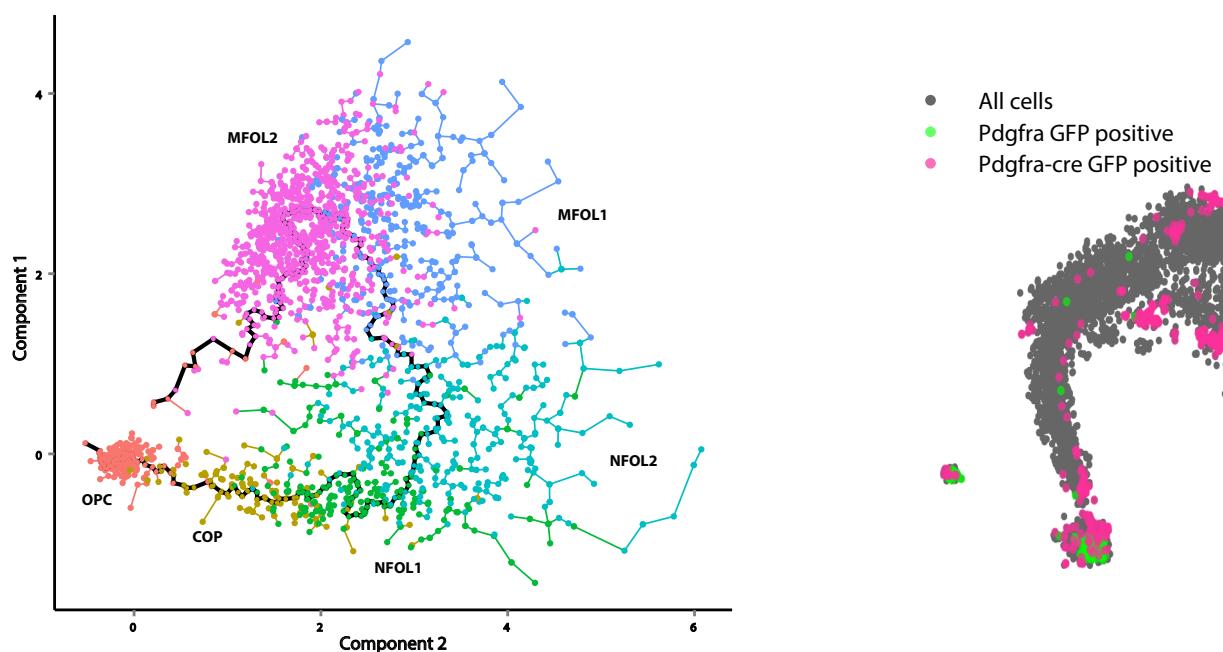


Fig. S1.

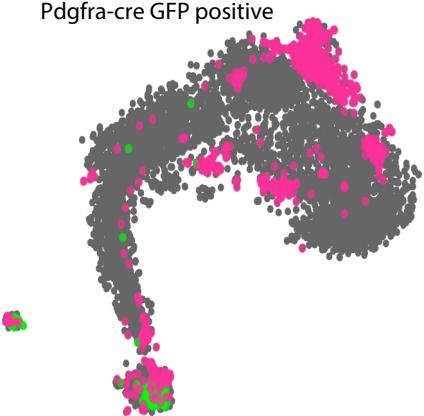
(A) Expression of canonical oligodendrocytes markers are overlaid on the tSNE map, gray = low expression and red = high expression. (B) Heatmap (green=low, yellow=high) show markers specific to each one of the resulted clusters. Upper panel show tissue of origin.

A

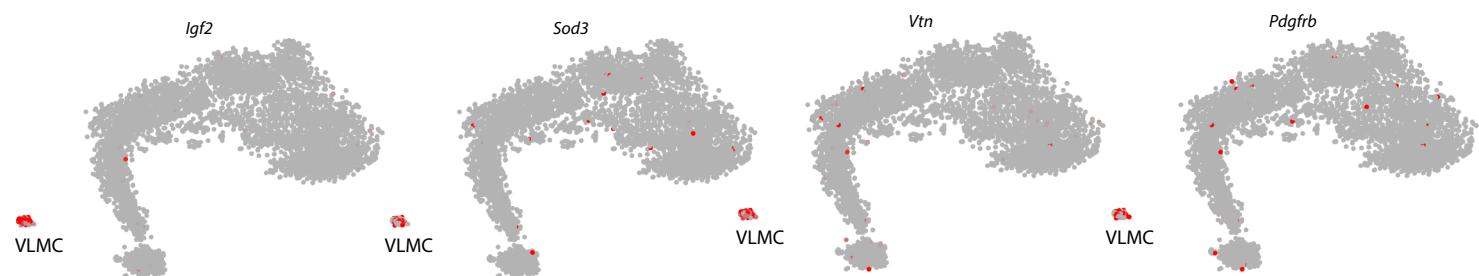


B

- All cells
- Pdgfra GFP positive
- Pdgfra-cre GFP positive



C



D

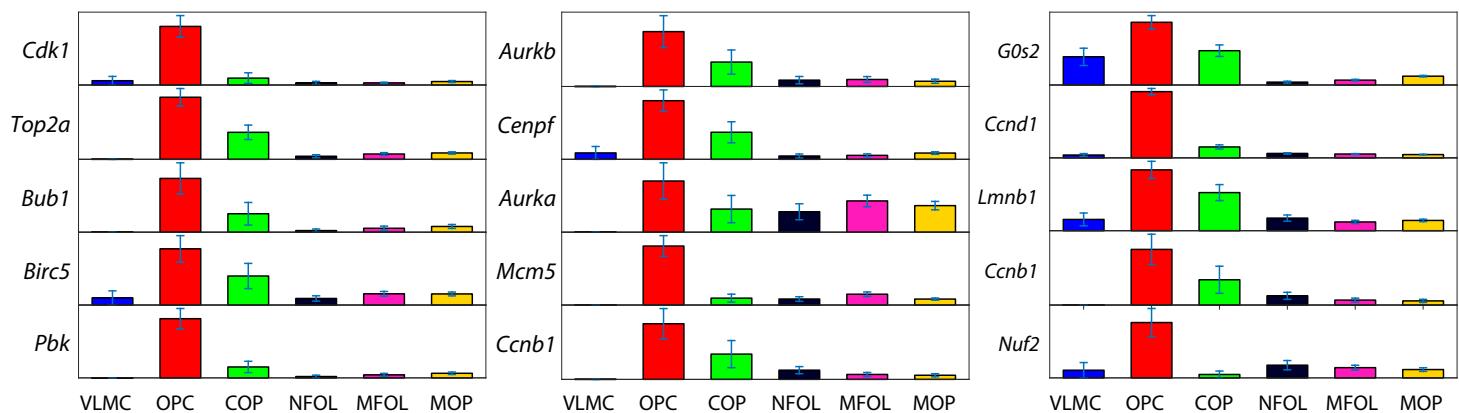


Fig. S2

(A) Unsupervised ordering of cells by expression state using Monocle based on independent component analysis (ICA) projection, using all cells except mature oligodendrocytes. (B) FACS-sorted cells, positive for GFP (by imaging) in both the *Pdgfra*-H2B-GFP (green) and the *Pdgfra*-Cre-RCE (pink) mice are overlaid on the tSNE map. Majority of GFP+ cells from *Pdgfra*-H2B-GFP mice were clustered into two populations (OPC and vessel and leptomeningeal cells). (C) Expression of pericytes markers is overlaid on the tSNE map showing strong enrichment in the vessel and leptomeningeal cell cluster. (D) Barplots showing the average expression of population (\pm s.e.m.) for a panel of cell-cycle genes suggest strong enrichment to the OPC cluster.

	Zhang et al 2014 database							
OPC markers	Astrocytes	Neurons	OPC	Newly formed OL	Myelinating OL	Microglia	Endothelial	
<i>Pdgfra</i>								
<i>Shc4</i>								
<i>Cspg4</i>								
<i>Gucy2f</i>								
<i>Smoc1</i>								
<i>Rlbp1</i>								
<i>Pcdh15</i>								
<i>Calcrl</i>								
<i>Kcnj16</i>								
<i>Emid1</i>								
<i>Tmem100</i>								
<i>Ednrb</i>								
<i>Zfp36l1</i>								
<i>Ccnb1</i>								
<i>Ska1</i>								
<i>Fabp7</i>								
<i>Pde6g</i>								

VLMC markers	Astrocytes	Neurons	OPC	Newly formed OL	Myelinating OL	Microglia	Endothelial
<i>Pdgfra</i>							
<i>Lum</i>							
<i>Fam180a</i>							
<i>Itih2</i>							
<i>Slc22a6</i>							
<i>Col3a1</i>							
<i>Dcn</i>							
<i>Col1a2</i>							
<i>Col1a1</i>							
<i>Mfap4</i>							
<i>Slc6a13</i>							
<i>Col3a1</i>							
<i>Igf2</i>							
<i>Pcolce</i>							
<i>Tbx18</i>							
<i>Nupr1</i>							
<i>Vtn</i>							
<i>Igfbp2</i>							
<i>Serpinh1</i>							

COP markers	Astrocytes	Neurons	OPC	Newly formed OL	Myelinating OL	Microglia	Endothelial
<i>Matn1</i>							
<i>Neu4</i>							
<i>Bmp4</i>							
<i>Sox6</i>							
<i>Gpr17</i>							
<i>Fyn</i>							
<i>Mpzl1</i>							

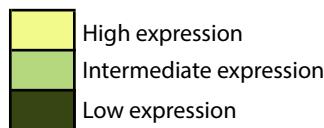


Fig. S3

Comparison of vascular and leptomeningeal cells, OPC and differentiation-committed oligodendrocyte precursor specific markers with transcriptomics database of Zhang *et al* 2014 (48). The top markers in the each population were applied to http://web.stanford.edu/group/barres_lab/brain_rnaseq.html, which features the database from (48). Depending on the level of expression, the genes were assigned as Low (basal expression with low Fpkm), Intermediate (between basal and higher Fpkm levels) or High (highest Fpkm levels) expression in the corresponding cell types. The results were then plotted in tables with Low expression depicted in dark green boxes, Intermediate in light green and High expression in yellow boxes.

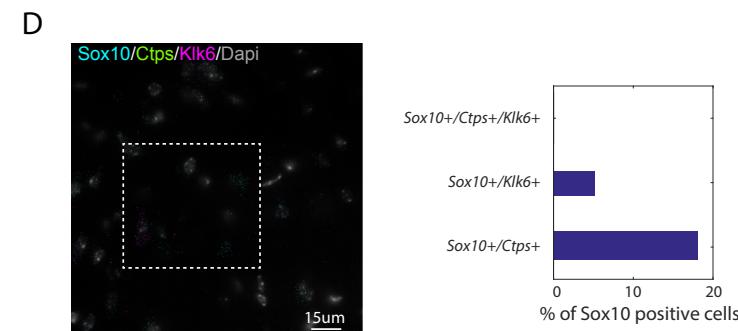
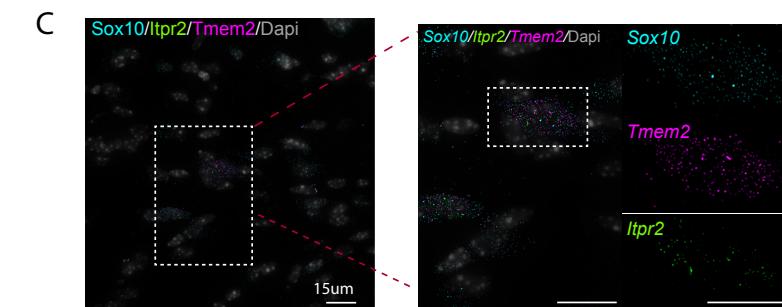
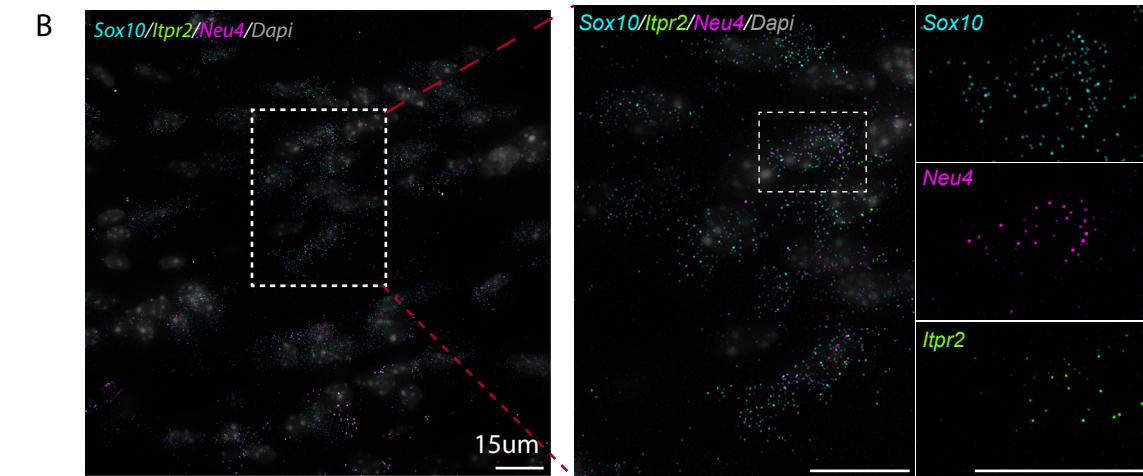
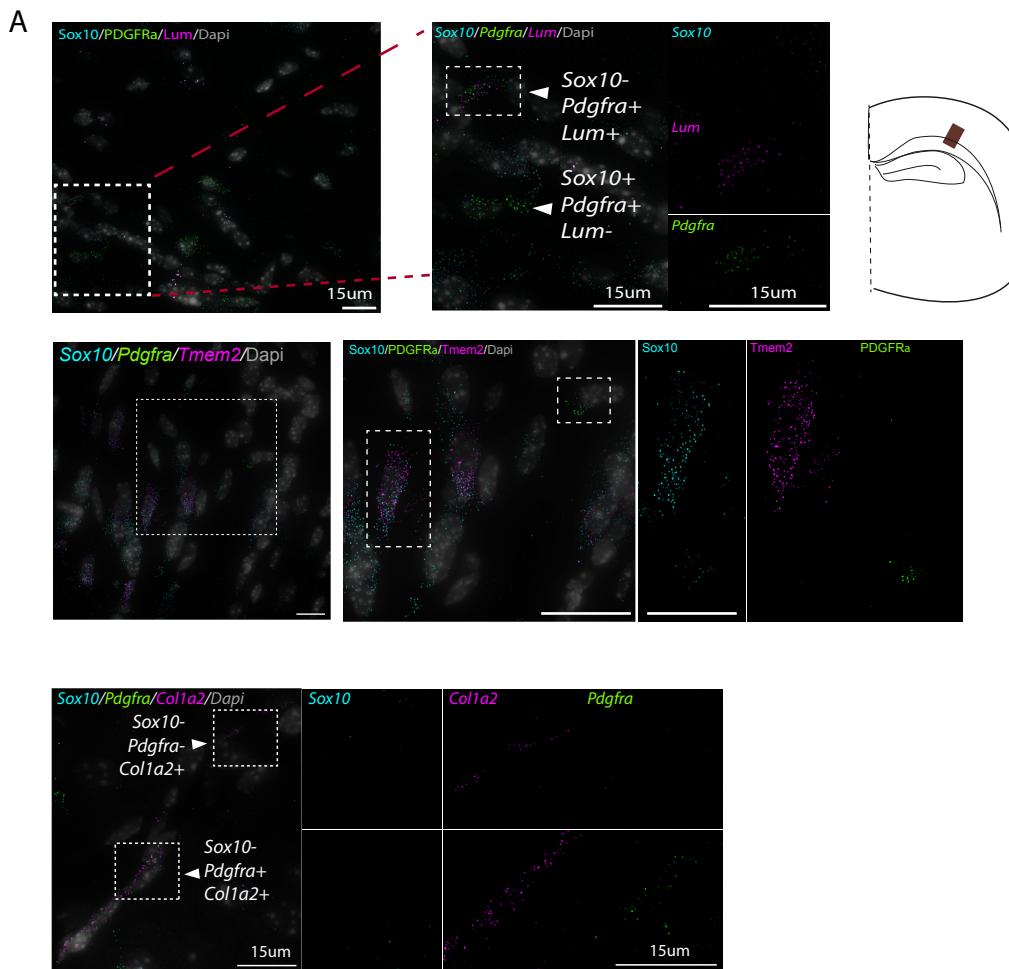


Fig. S4

Validation of subpopulation specific markers using smFISH. (A) Targeting *Sox10* with *Pdgfra* and *Lum* (vascular and leptomeningeal cells) markers, *Tmem2* (newly-formed oligodendrocytes specific marker) and *Colla1* (vascular and leptomeningeal cells specific marker). Cells with co-expression of *Pdgfra* and *Lum* are negative for *Sox10* (dashed rectangle) while cells which co-express *Pdgfra* and *Sox10* are negative for *Lum* hence vascular and leptomeningeal cells and OPC are distinct *Pdgfra*⁺ populations. Cells expressing *Pdgfra* are not colocalized with *Tmem2*⁺ cells, showing that OPC and newly formed oligodendrocytes are distinct populations. Cells co-expressing *Pdgfra* and *Colla1* are negative for *Sox10* (dashed rectangle) and lie on a blood vessel, supporting that they are not from the oligodendrocyte lineage and their relation with pericytes. (B) Targeting *Sox10*, *Itpr2* (early stages of oligodendrocytes differentiation marker) and *Neu4* (committed precursors specific marker). (C) Targeting *Sox10*, *Itpr2* and *Tmem2* (specific newly formed oligodendrocytes marker). (D) Targeting *Sox10*, *Ctps* (myelin forming oligodendrocytes marker) and *Klk6* (mature oligodendrocytes 2 marker), (wide view from Figure 2). Quantification of myelin forming oligodendrocytes (*Sox10*⁺/*Ctps*⁺) and population 2 of mature oligodendrocytes (*Sox10*⁺/*Klk6*⁺) cells. No cells expressed both *Ctps* and *Klk6* showing they are distinct populations.

Junction 1: VLMC vs OL lineage groups

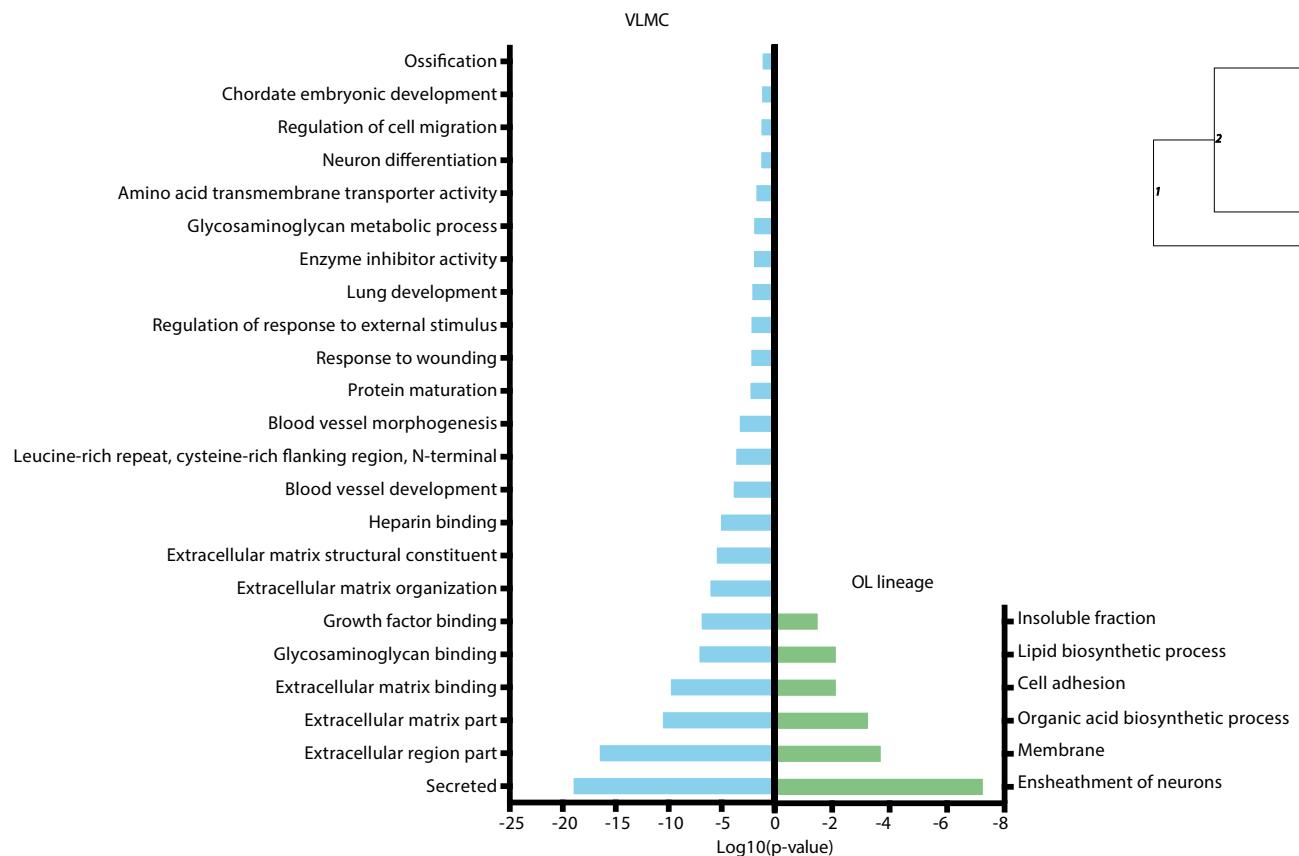
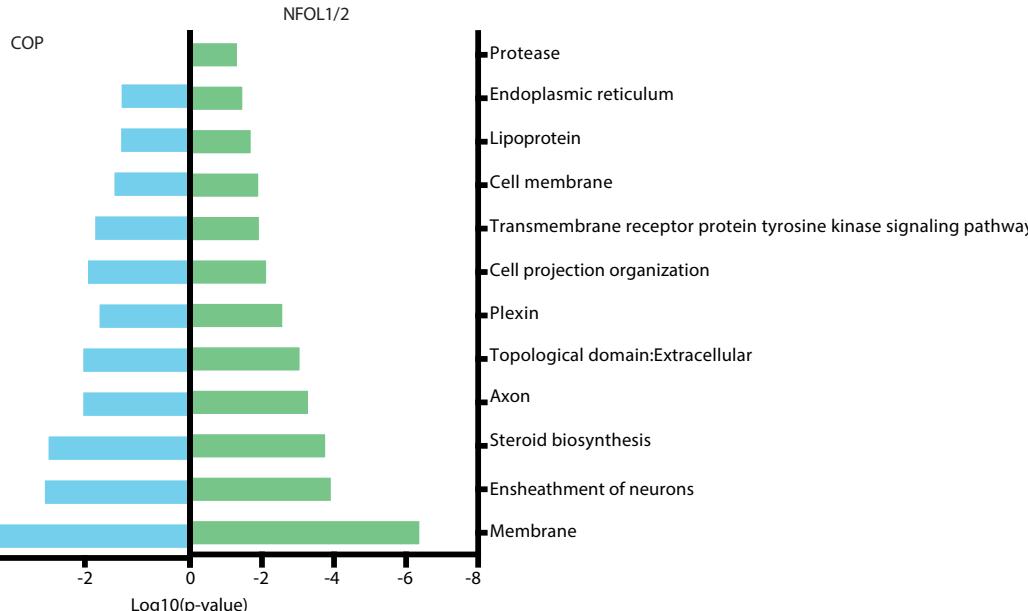
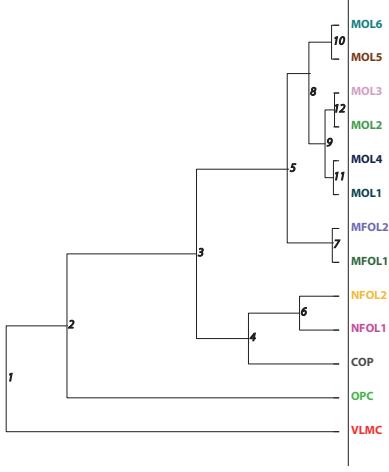


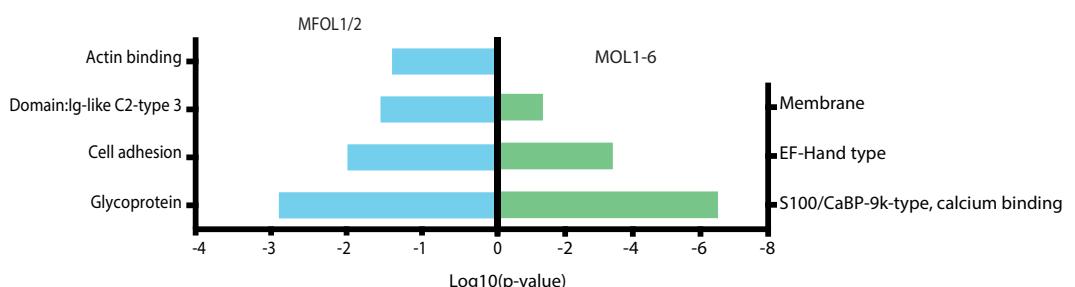
Fig. S5

GO analysis of top 50 regulated genes in each dendrogram comparison. All genes were ranked by the difference in average expression in population A vs B (related to dendrogram obtained by linkage clustering). The list of 50 top genes in each branch was then applied to DAVID v6.7 and plots of the most significant categories were plotted (depending on Log10 (p-value)). Graphs of junction 1: Vascular and leptomeningeal cells vs oligodendrocyte lineage (all remaining cell types); junction 2: OPC vs other differentiated oligodendrocytes and junction 3: early stages (committed oligodendrocyte precursors and newly-formed oligodendrocytes 1/2) vs intermediate and late stages (myelin-forming and mature oligodendrocytes).

Junction 4: COP vs NFOL1/2



Junction 5: MFOL1/2 vs MOL1-6



Junction 6: NFOL1 vs NFOL2

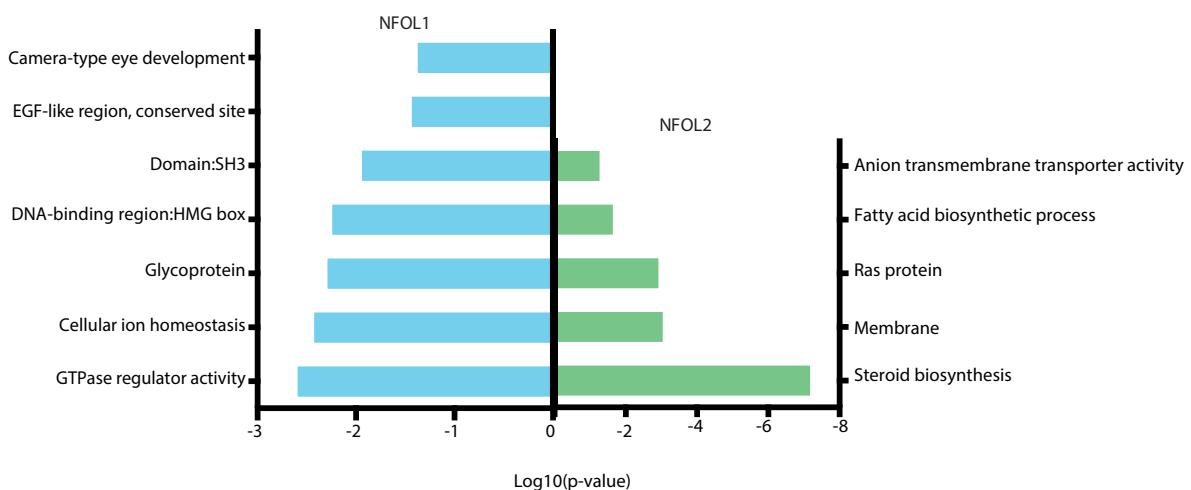
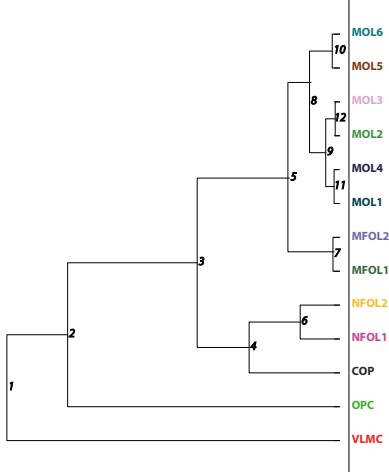
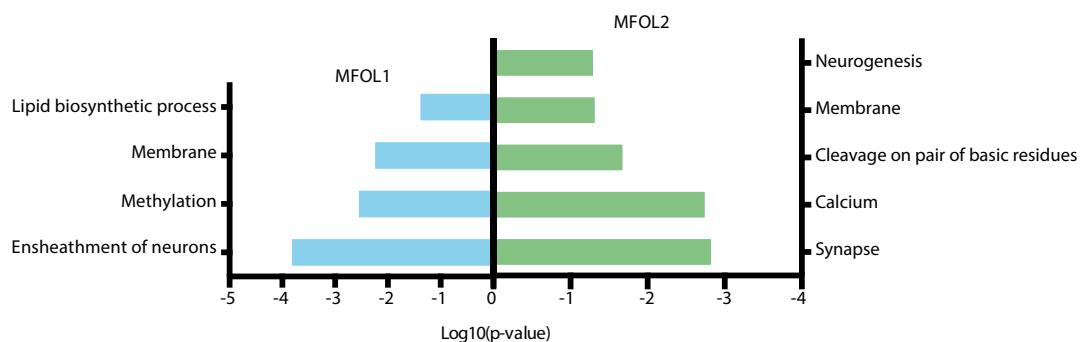


Fig. S6

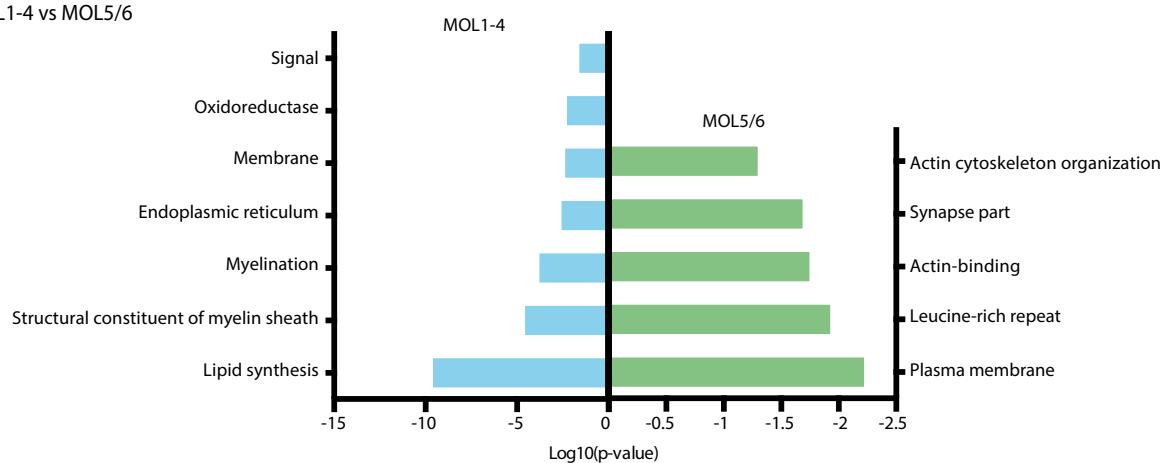
GO analysis of top 50 regulated genes in each dendrogram comparison. Obtained as explained in Figure S5 legend. Graphs of junction 4: committed oligodendrocyte precursors vs newly-formed oligodendrocytes 1/2; junction 5: Myelin-forming oligodendrocytes 1/2 vs other mature oligodendrocytes 1-6, junction 6: newly -formed oligodendrocytes 1 vs 2.



Junction 7: MFOL1 vs MFOL2



Junction 8: MOL1-4 vs MOL5/6



Junction 9: MOL1/4 vs MOL2/3

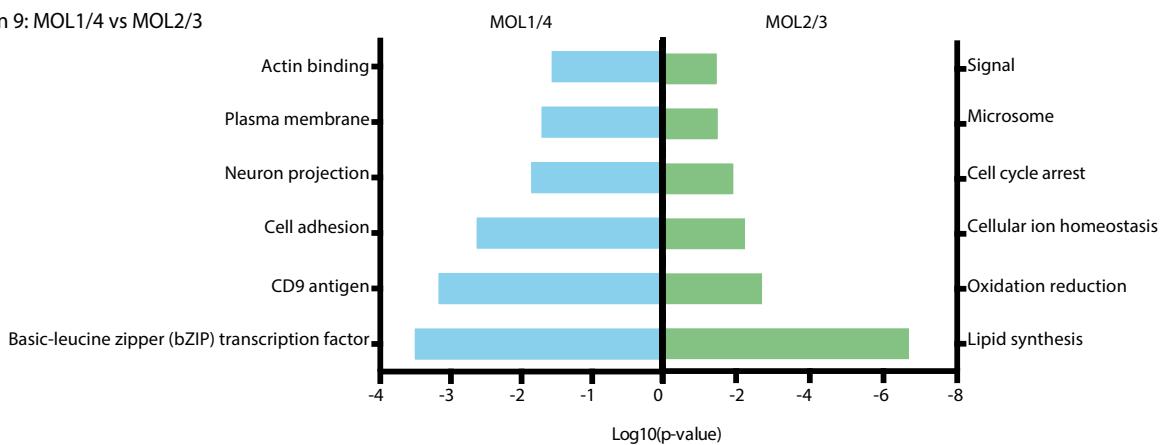
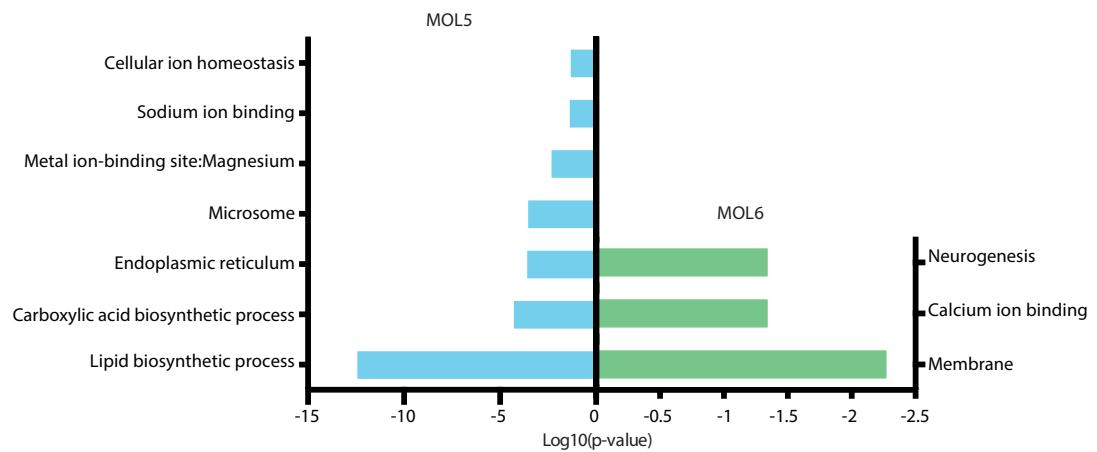
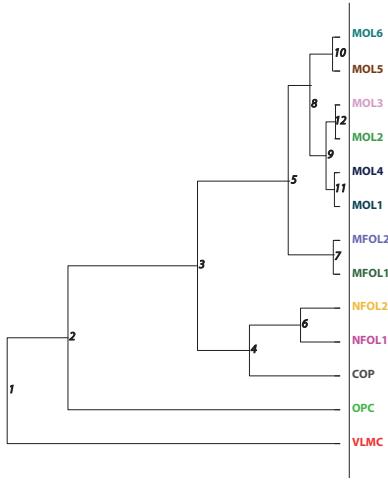


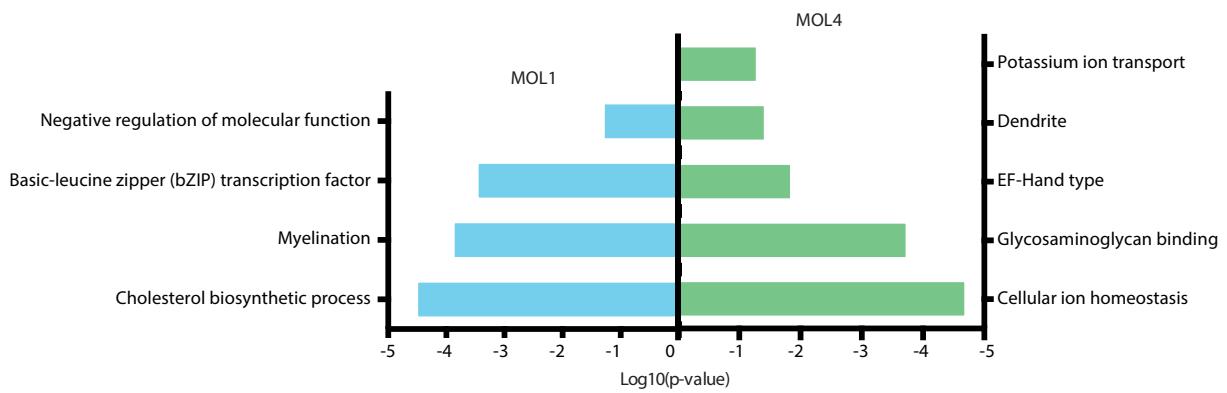
Fig. S7

GO analysis of top 50 regulated genes in each dendrogram comparison. Obtained as explained in Figure S5 legend. Graphs of junction 7: Myelin-forming oligodendrocytes 1 vs 2, junction 8: Mature oligodendrocytes 1-4 vs 5/6 and junction 9: mature oligodendrocytes 1/4 vs 2/3

Junction 10: MOL5 vs MOL6



Junction 11: MOL1 vs MOL4



Junction 12: MOL2 vs MOL3

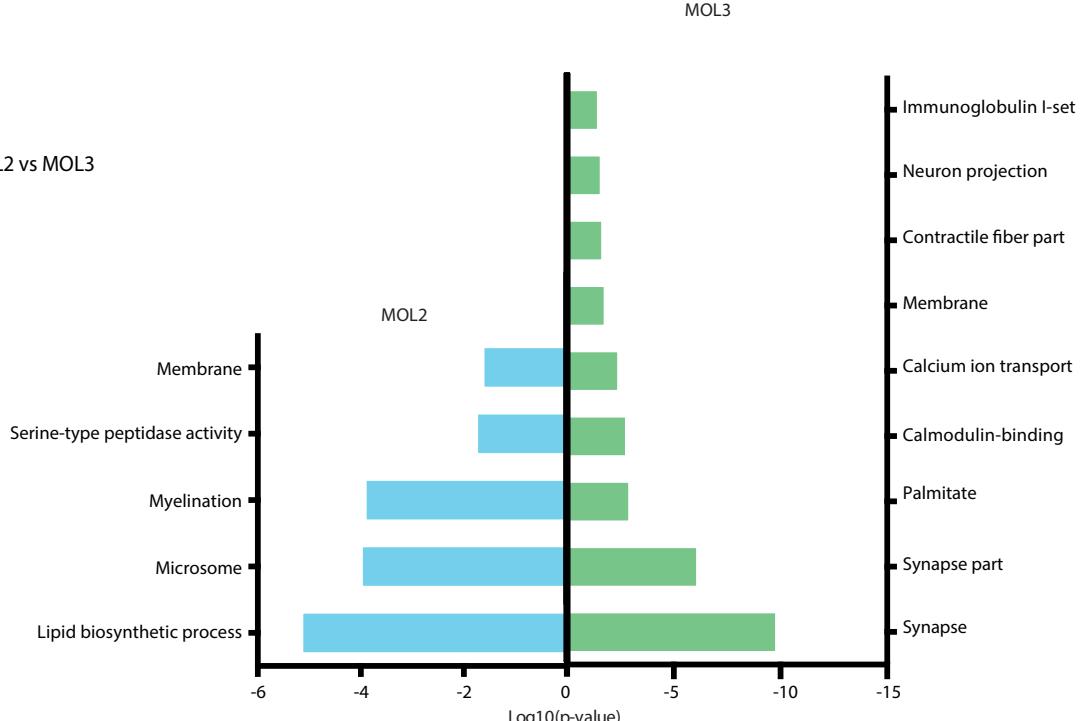


Fig. S8

GO analysis of top 50 regulated genes in each dendrogram comparison. Obtained as explained in Figure S5 legend. Graphs of junction 10: mature oligodendrocytes 5 vs 6, junction 11: mature oligodendrocytes 1 vs 4 and junction 12: mature oligodendrocytes 2 vs MOL3

TF=□

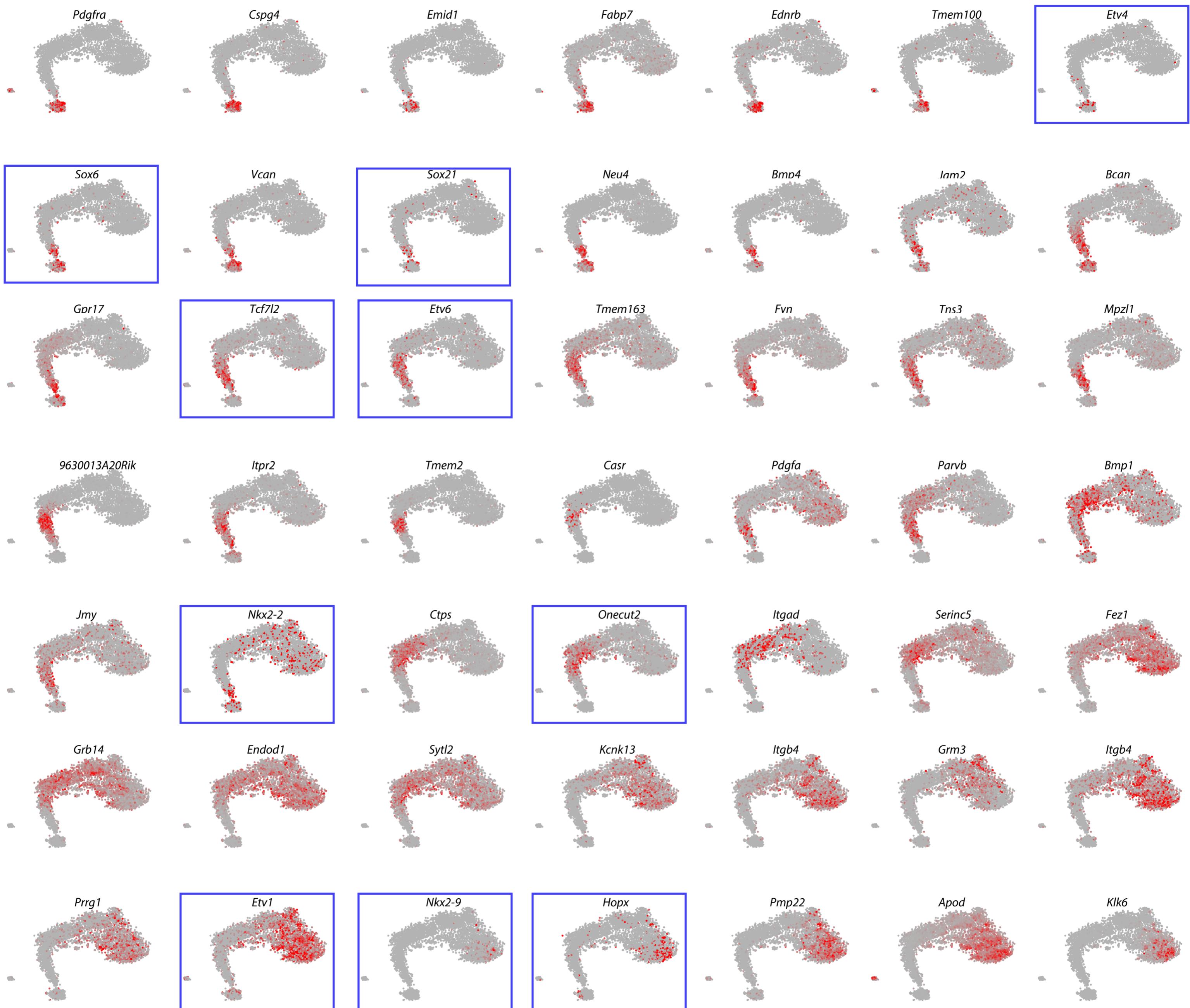
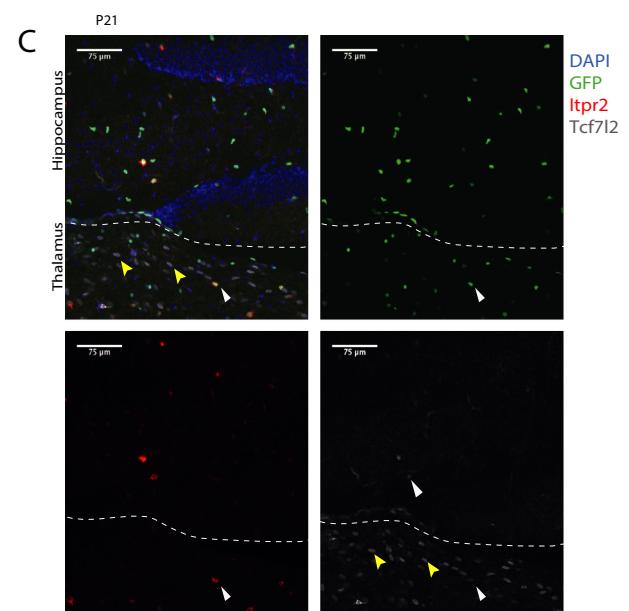
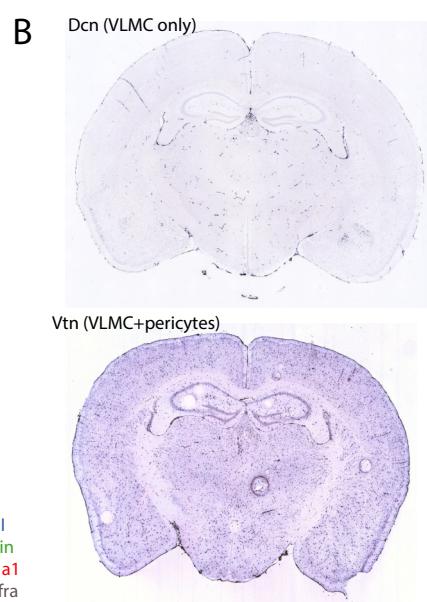
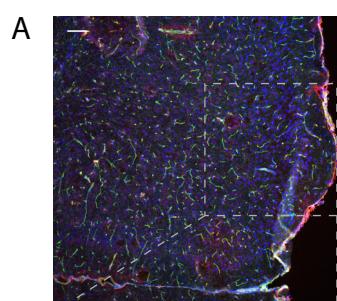


Fig. S9

Expression of stage specific genes along the maturation process overlaid on the tSNE map (gray = low expression and red = high expression). Genes are ordered from specificity in OPC to mature oligodendrocytes. Transcription factors mentioned in the main text are highlighted with a blue square.



D

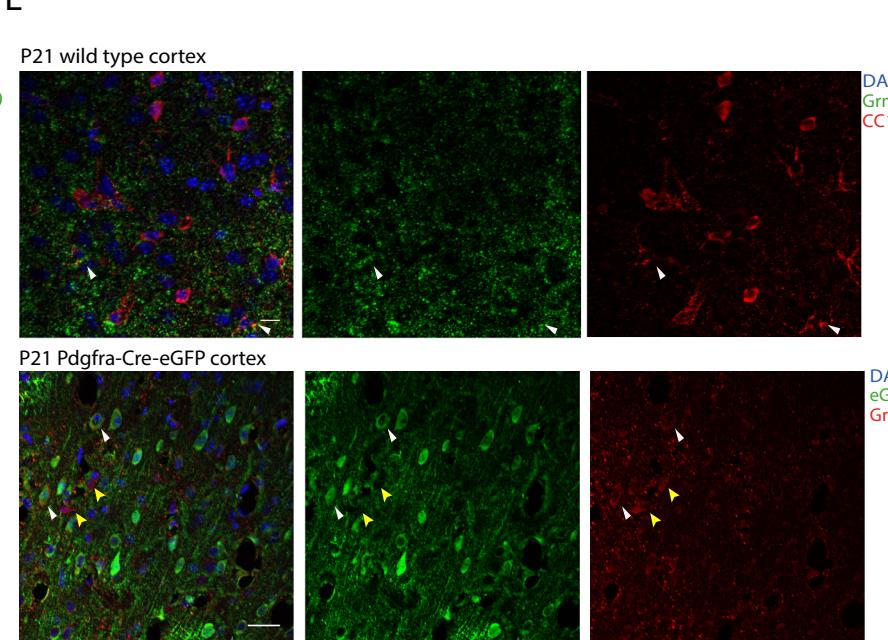
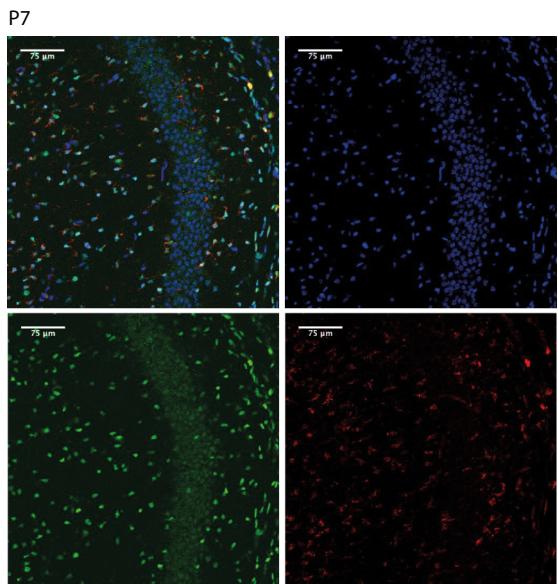


Fig. S10

(A) Staining of Col1a1 (VLMC marker), Pdgfra and Tomato lectin (marker of blood vessels) in P21 brain, shows the presence of VLMC in the meninges. Scale bar 25 μ m (B) Allen brain Atlas (<http://www.brain-map.org/>) coronal sections with ISH of Dcn (marker of VLMC) and Vtn (VLMC and pericytes) show that VLMCs are mainly in meninges and large vessels. (C) Tcf7L2 (grey) was not specific to GFP+ cells (from *Pdgfra*-H2B-GFP mice) in specific areas of the brain, such as the thalamus. (D) Co-staining of Sox10 (green) with Itpr2 (red) in P7 CA1 and Hippocampus. Images of the 3 channels combined and in separate are shown to highlight the high co-localization of Itpr2 in Sox10+ cells at this time point. (E) Staining of Grm3 was performed in sections of P21 wild type cortex, together with CC1, which labels mature oligodendrocytes and in P21 *Pdgfra*-Cre-RCE cortex (which stain the entire oligodendrocyte lineage). Images of the 3 channels combined and in separate are shown to highlight the co-localization of Grm3 in some of the CC1/GFP cells (white arrowheads), although also labelling other cell types, probably neurons (yellow arrowheads). Scale bar– 25 μ m.

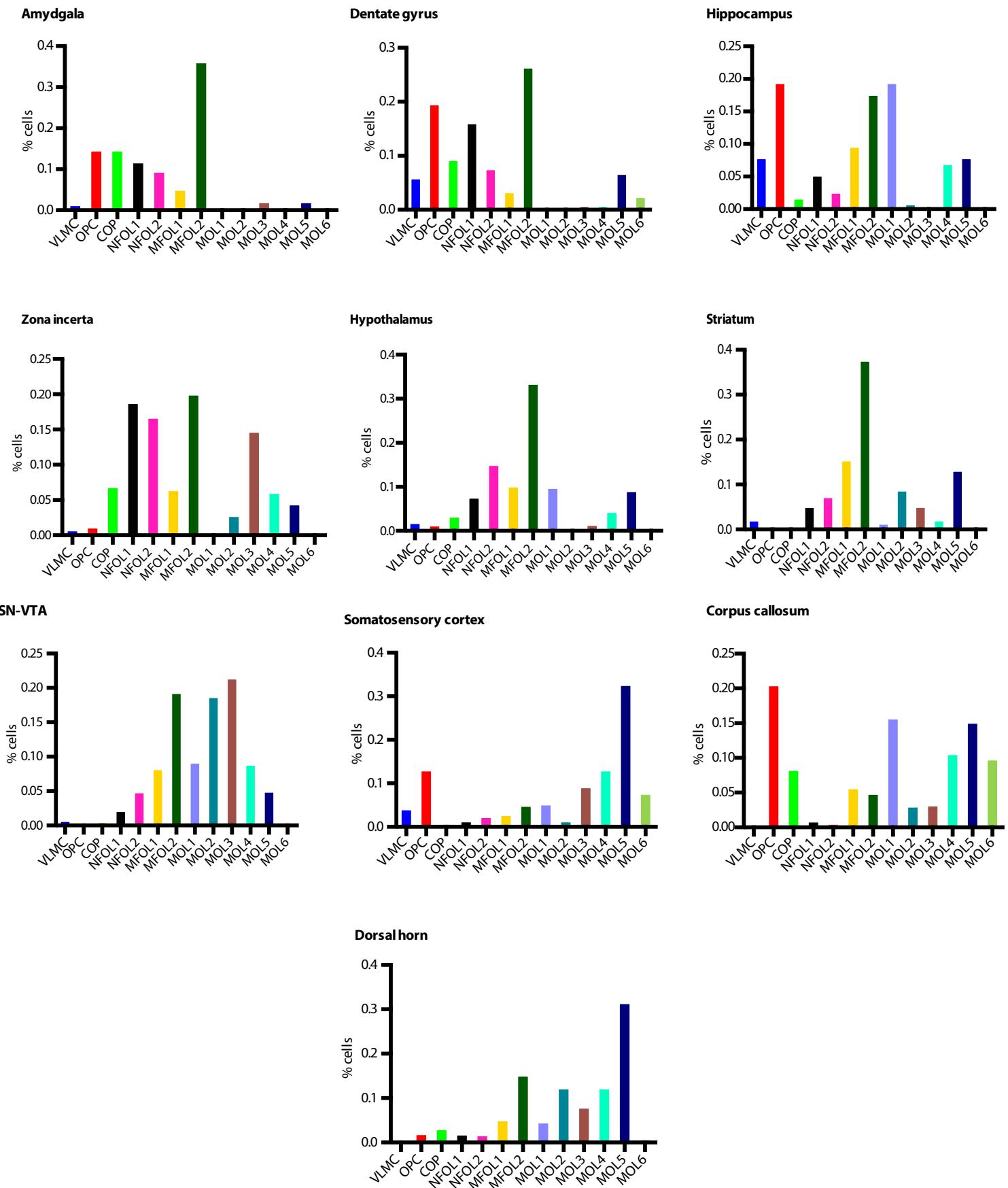


Fig. S11

Percentage of cells from each cell type along the different CNS tissues. The percentage of cells contributing to each subgroup in a specific CNS region was plotted.

Marques and Zeisel et al, Figure S12

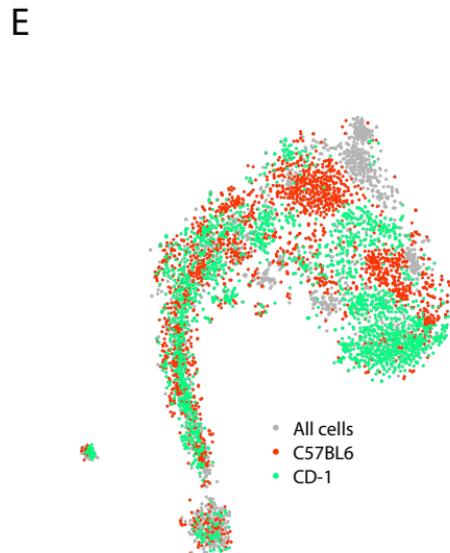
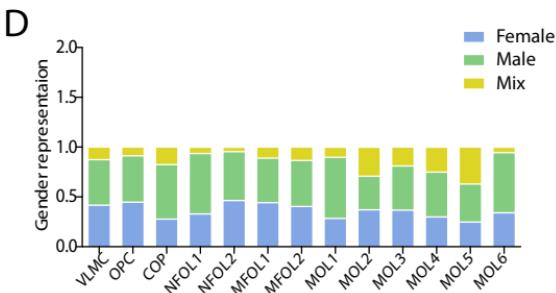
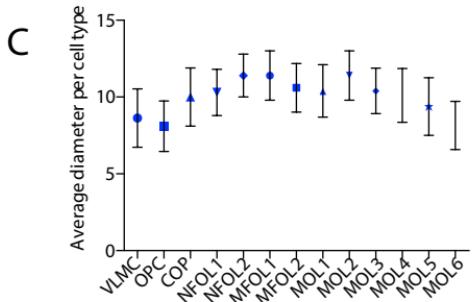
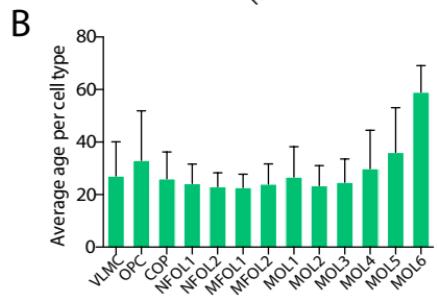
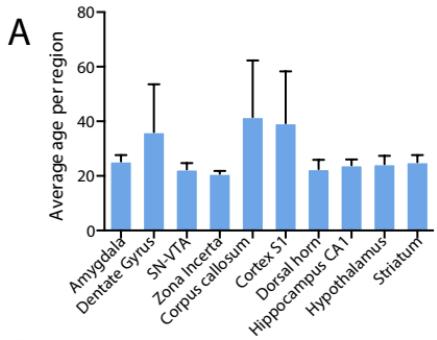


Fig. S12

Age, gender and cell size parameters across the dataset. Average of the age of the mice used per CNS regions(A) and per cell type (B) with standard deviation. (C) Average and standard deviation of the diameter of each cell stage.. (D) The fraction of males and females that contributed to each cell type. Blue represents females, green being males and yellow being mixed or with uncertain gender. (E) Representation different genetic backgrounds used in the study (wild type C57BL6 and wild type CD-1) overlaid on the tSNE map. Cells from transgenic mice lines are in grey.

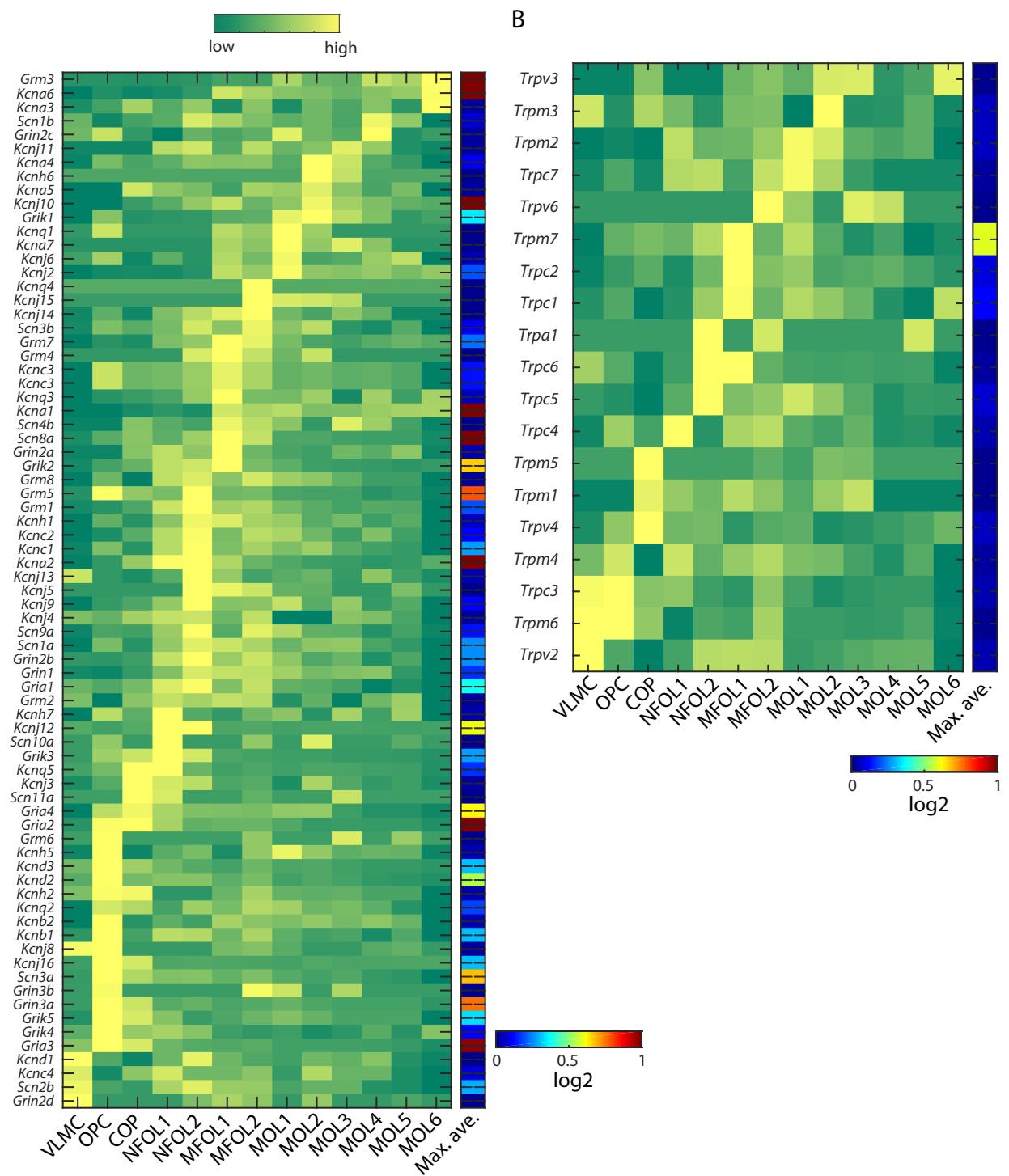


Fig. S13

Heatmap with group average expression profile of ionotropic and metabotropic glutamate receptors and other ions channels (A), and TRP channels (B) in the identified oligodendrocyte populations (green=low, yellow=high). Each row is centered normalized (log2 z transform). Bars on the right (blue (low) to red (high) scale) represents the maximum of the average expression along the groups in absolute values. Genes with a high (yellow) expression in the heat map might have a low maximum of the average expression along the groups (blue), due to scattered expression (expression only in a subset of single cells).

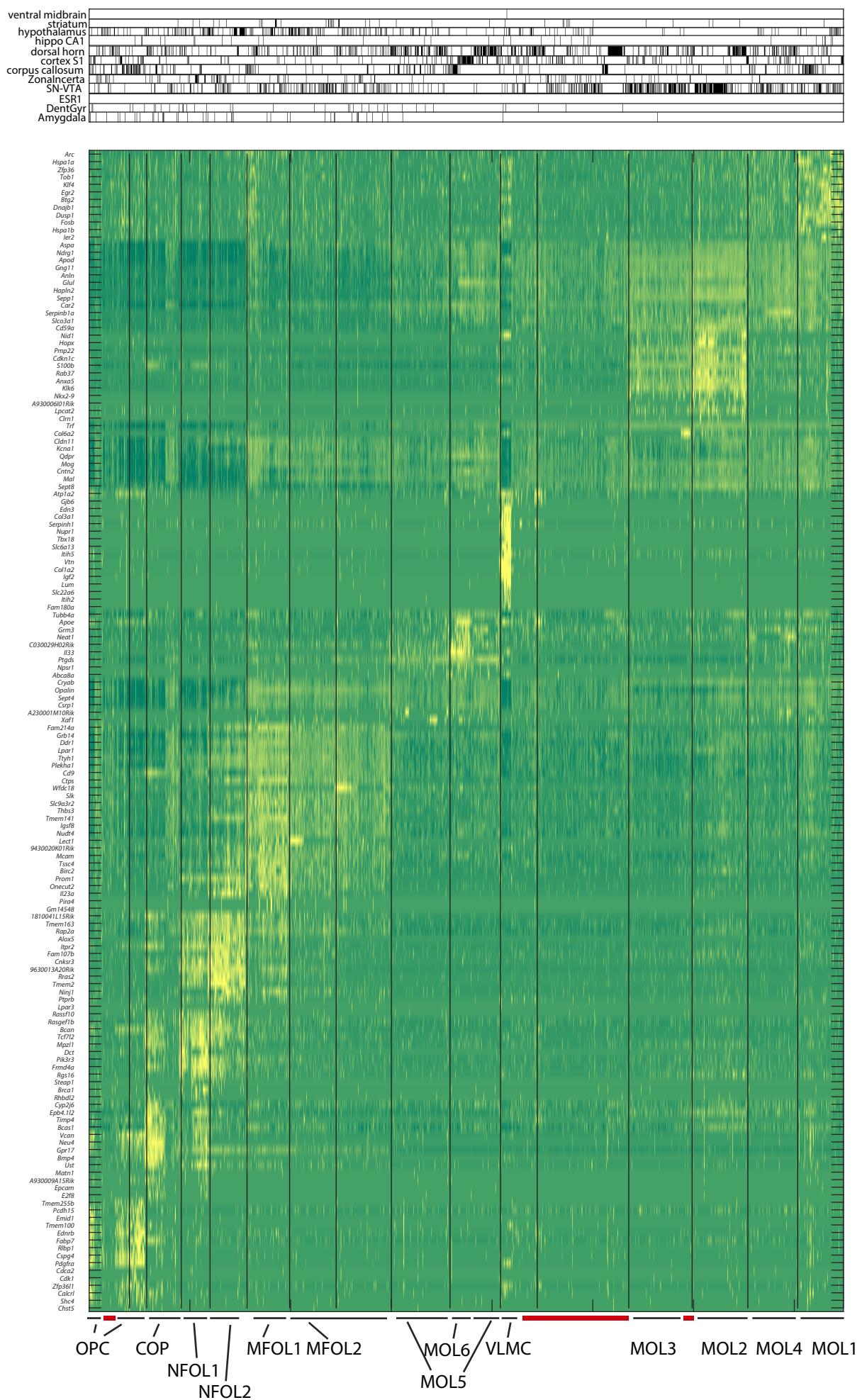


Fig. S14

Heatmap (green=low, yellow=high) show markers specific to each one of the clusters after running BackSpin (see methods). The scheme of cluster merging and excluded cells shown on the bottom (red bar = excluded cells).

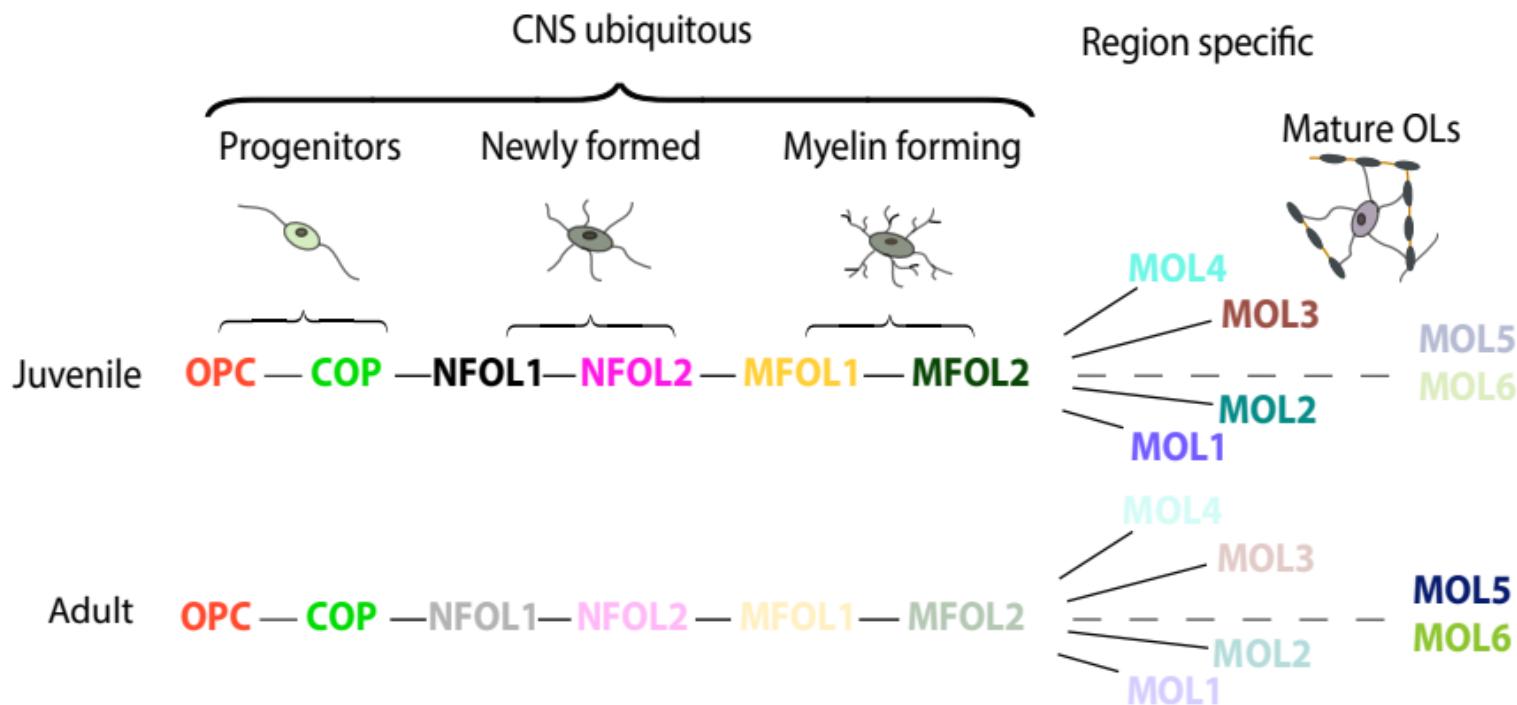


Fig. S15

Model of oligodendrocyte lineage progression, based on single cell RNA-Seq data: Oligodendrocyte maturation process is sequential, continuous and uniform along CNS until myelin-forming cells, while mature populations show heterogeneity and region specificity. The subpopulations represented in the adult tissue are mainly OPCs and MOL5/6 (MOL6 unique for P60 stage), being the intermediate population almost non-existent.

Additional Data Table S1- List of 50 up-regulated genes in each branch of the dendrogram.

Additional Data Table S2 - Genes that contribute to the enriched functional categories in each dendrogram group comparison.

Additional Data Table S3- – List of mice strains used in the study by region, depicting also the background and if FACS sorted.