Injury and loss of oligodendrocytes can cause demyelinating diseases such as multiple sclerosis. To improve our understanding of human oligodendrocyte development, which could facilitate development of remyelination-based treatment strategies, here we describe time-course single-cell-transcriptomic analysis of developing human stem cell-derived oligodendrocyte-lineage-cells (hOLLCs). The study includes hOLLCs derived from both genome engineered embryonic stem cell (ESC) reporter cells containing an Identification-and-Purification tag driven by the endogenous PDGFRα promoter and from unmodified induced pluripotent (iPS) cells. Our analysis uncovers substantial transcriptional heterogeneity of PDGFRα-lineage hOLLCs. We discover sub-populations of human oligodendrocyte progenitor cells (hOPCs) including a potential cytokine-responsive hOPC subset, and identify candidate regulatory genes/networks that define the identity of these sub-populations. Pseudotime trajectory analysis defines developmental pathways of oligodendrocytes vs astrocytes from PDGFRα-expressing hOPCs and predicts differentially expressed genes between the two lineages. In addition, pathway enrichment analysis followed by pharmacological intervention of these pathways confirm that mTOR and cholesterol biosynthesis signaling pathways are involved in maturation of oligodendrocytes from hOPCs.
**Results**

**Generation of an OPC differentiation and purification stem cell reporter line.** Several protocols for differentiation of OLLCs from hPSCs have been reported\(^\text{12-14,20-22}\). The hOPCs from such differentiating cultures can be purified using antibodies against endogenous O4 or A2B5 surface antigens or against PDGFR\(\alpha\), an OPC-specific surface protein. However, the majority of the O4\(^+\) cells represent post-mitotic immature oligodendrocytes and the A2B5\(^+\) cells consist of a heterogeneous population of glial restricted cells and developing neurons\(^3\). More importantly, since O4 and A2B5 are ganglioside epitopes, there is no easy way to genetically label and track the cells that express these antigens. Since we wanted to monitor the OL differentiation and study them starting at the early progenitor cell stage, we established a platform that enabled us to detect and purify cells expressing PDGFR\(\alpha\), a well-established marker for OPCs and pre-OL cells\(^\text{11,20,23}\). We created a reporter hESC line in which the identification-and-purification (IAP) sequence\(^\text{24}\) was engineered to be expressed under the control of PDGFR\(\alpha\). The IAP tag consists of a tdTomato fluorescent marker and a mouse cell-surface protein, Thy1.2, separated from each other and from the endogenous PDGFR\(\alpha\) gene product by the “ribosome-skipping” 2A peptide\(^\text{24}\) (P2A-tdTomato-P2A-Thy1.2) (Fig. 1a). We have previously demonstrated the efficacy of the IAP reporter system for detecting and purifying cells of interest from hESC-derived heterogeneous cell populations, and have shown that P2A functions, as reported\(^\text{25-26}\), to separate the translated gene products so that the endogenous gene remains functionally intact while the fluorescent reporter is cytoplasmic and Thy1.2 is present on the cell membrane\(^\text{24}\).

To generate the PDGFR\(\alpha\)-P2A-tdTomato-P2A-Thy1.2 hESC reporter (PD-TT), the ESC line WA09 (H9) was engineered to integrate the IAP tag at the 3’-end of the PDGFR\(\alpha\) gene using CRISPR-based genome editing (Fig. 1a–c). Using transient antibiotic selection\(^\text{27}\), we achieved ~20% homozygous knock-in of the reporter sequence (Supplementary Fig. 1a, b). Sequencing confirmed the absence of mutations at the predicted five most likely off-target genomic locations (Supplementary Fig. 1f), and karyotype analysis performed at different passage numbers showed no abnormality in the early passages (Supplementary Fig. 1c). However, we found isochromosome duplication of the long arm of chromosome 1 in 50% of the clones analyzed by G-banding in the cells after passage #10 (Supplementary Fig. 1c–e). This change is one of the most common karyotypic abnormalities found in hESCs, comprising 10–25% of the total hESCs with aberrations\(^\text{28,29}\).

**Differentiation and purification of PD-TT-derived OPs.** We differentiated the hESC PD-TT reporter line into OPCs, following the protocol of Douvaras and Fossati\(^\text{14,30}\) (Supplementary Fig. 1g). Analogous to the timing of initial PDGFR\(\alpha\) mRNA expression, small clusters of tdTomato\(^+\) cells were visible in the differentiation culture as early as day 8. However, morphologically bipolar, individual, tdTomato\(^+\) OPCs were not visible until ~day 45, at which time mRNA levels of PDGFR\(\alpha\) are increased ~700-fold compared to undifferentiated PD-TT cells (Supplementary Fig. 1h, i and Supplementary Movie 1). By day 60, numerous tdTomato\(^+\) cells were seen migrating out from neurospheres grown on poly-\(\gamma\)-ornithine/laminin-coated plates (Supplementary Fig. 1h and Supplementary Movie 2). By day 80, when grown in mitogen-free glial media\(^\text{30}\), ~25% of the total cells in the differentiating cultures are tdTomato\(^+\) OPCs (Fig. 2a, c).

Since the differentiated reporter OPCs also express the mouse Thy1.2 surface tag, these cells can be immunopurified via anti-Thy1.2 microbeads and magnetic-activated cell sorting (MACS)
Fig. 2b. Of note, the antibodies against mouse Thy1.2 used in the purification are species-specific and do not react against human Thy1.24,31, and immunopurification using the Thy1.2 surface tag is gentler to the cells and is more amenable to purifying large numbers of cells than fluorescent activated cell sorting (FACS)24. Using MACS purification, we routinely obtain an ~90% pure population of tdTomato+ PDGFRα+ expressing cells (Fig. 2c, d). When the purified cells are re-plated on a laminin-coated surface, they show bipolar morphology, a characteristic feature of OPCs (Fig. 2d). As they mature, within two weeks in the mitogen withdrawn glial media, tdTomato expression in the cells is reduced and multiple branches and fine cell processes are formed resembling either astrocyte or oligodendrocyte cells (Supplementary Fig. 2a).

Characterization of the PD-TT cells after OPC differentiation.

To further characterize the PDGFRα-tdTomato-Thy1.2 expressing cells, at day 80 of differentiation, Thy1.2 antibody binding cells were separated from the non-binding cells by MACS purification. Flow analysis showed ~90% of the binding cells were tdTomato+ (Fig. 2c). The expression of various OPC markers between the purified and unbound cells was analyzed by quantitative reverse-transcription PCR (qPCR) (Fig. 2e). The qPCR expression data were normalized relative to PD-TT undifferentiated stem cells. The purified cells showed significantly enhanced expression of PDGFRα and other OPC markers, including SOX10, OLG2, CSPG4 (NG2), and ENPP6. Transcripts for MBP, an oligodendrocyte marker gene, was enriched in the purified cells compared to the unbound cells whereas GFAP, a gene that is expressed at high levels in astrocytes, was partially enriched in the unbound population (Fig. 2e). Immunofluorescence showed that the majority of tdTomato+ cells express NKX2.2, OLIG2, and SOX10 (Fig. 2g and Supplementary Fig. S2b), and 80% of the tdTomato+ cells are also PDGFRα+ (Fig. 2f and Supplementary Fig. 3a, b). Furthermore, higher resolution images of cells stained for PDGFRα showed that PDGFRα localizes to the cell membrane while the tdTomato is cytoplasmic (Fig. 3a and Supplementary Fig. 2c), which indicates that the PDGFRα protein product is effectively separated from the reporter proteins.

To further examine whether the differentiating reporter cells are representative of the OL progenitors, we looked at the O4 antigen expression in the tdTomato-expressing cells. At day 85, ~20% of the total cells in a differentiating culture were O4+ and 70% of the O4+ cells expressed tdTomato (Fig. 3b, c). Only ~30% of the A2B5+ cells were tdTomato+ at this stage (Supplementary Fig. 3c). When we purified the O4+ cells from day 115 differentiating culture using O4 microbeads, we found ~87% overlap between O4+ and tdTomato+ cells (Fig. 3d). These results are comparable to the amount of overlap between PDGFRα+ (CD140a+) and O4+ cells in primary and hPSC-derived human OPCs11,21, and shows that expression of the reporter occurs in both early progenitors as well as pre-OL cells and thus can be used to purify cells at various stages of the oligodendrocyte lineage.

Fig. 1 Generation of PDGFRα reporter ES cell line. a Schematic diagram of IAP reporter knock-in into the PDGFRα locus using CRISPR-Cas9 genome editing. A plasmid containing spCas9 sequence and a guide sequence targeting the stop codon of the PDGFRα gene, and a separate “donor” plasmid containing reporter sequence flanked by 1 kb homology arms were transfected into H9 ES cells. Following single-cell passaging, PCR-based genotyping was performed on individual clones using a primer set that spans the reporter sequence (one primer at the homology arm and the other outside the homology arm). b A representative PCR gel image that differentiates between WT, heterozygous, and homozygous knock-in clones. A full scan of the gel is included in Supplementary Fig. S1a. The PCR-based genotyping was independently repeated five times with a similar result. c Sanger sequencing of the knock-in band to confirm insertion of the reporter at the correct location.
When purified tdTomato+ cells were plated on either laminin-coated surface or plates containing electrospun nanofibers for another 3 weeks, they continued to differentiate into mature OLs. The OL cultures were stained with an antibody against MBP and MBP+ cells displayed the distinct, typical branched morphology of mature OLs (Fig. 3e). The cells cultured on electrospun nanofibers not only developed processes but the processes were aligned along the nanofibers, and appear to myelinate them (Fig. 3f). In addition, the MACS purified cells can be cryopreserved and revived with >80% viability. Upon revival,
Fig. 2 Differentiation and purification of PD-TT hESC reporter cells. a Eighty-days-old OPC culture expressing tdTomato driven by the endogenous PDGFRα promoter. Differentiation of the PD-TT reporter cells was performed ten times with similar results. Scale bar: 400 μm. b Schematic figure of MACS-based immunopurification of reporter cells using anti-Thy1.2 microbeads and a magnetic column. c Flow analysis indicating 23.1% of the total cells in differentiating culture are tdTomato+ at day 78, which is enriched to 90.1% after MACS purification. PE in X axis represents PDGFRα-tdTomato. A gate for flow analysis was set up using non-genome-edited hES cells differentiated to day 95 (left panel). Gating strategy is further detailed in Supplementary Fig. S1 and is also described in the “Methods” section. d Majority of the Thy1.2 immunopurified cells are tdTomato+ and have the OPC-like bipolar morphology. Scale bar: 200 μm. MACS-based purification and culture of the purified reporter OPCs were performed 10 times with similar result. e qPCR analysis shows enrichment of OPC markers in the MACS-based tdTomato enriched population vs the flow through. Two biological and three technical replicates each were used for qPCR analysis. Data are presented as mean ± SEM. Biological and technical replicates are distinguished by solid vs open symbols used for each data point. Source data for the qPCR are provided as a Source Data file. f−g Immunohistochemistry demonstrating that the MACS purified tdTomato+ cells express PDGFRα (f) and the OPC markers NKKX2.2 and SOX10 (g). Scale bar: 200 μm. Immunohistochemistry was independently repeated twice with similar results.

they maintain the capacity to mature into MBP+ OLs (Supplementary Fig. 3d, e). Therefore, these reporter cells provide the flexibility to collect large numbers of cells for studies that require a large amount of material (i.e., biochemical studies, screening applications, etc.).

Single-cell transcriptome analysis of PDGFRα+ reporter OLs. To better understand the gene expression nuances associated with human OPC/OL differentiation, we applied the microfluidic-based “Drop-seq” strategy16 to capture the transcriptome of differentiating hOLLCs at the single-cell level. PD-TT reporter cells at three timepoints spanning the in vitro differentiation process (days 77, 89, and 104) were separately MACS purified to ~90% PDGFRα-tdTomato purity, single cells captured, and their transcriptomes determined. After quality control (Supplementary Fig. 4a, Supplementary Data 5, detail in “Methods” section), a combined total of 3271 cells were used for further analysis. Seurat-based unsupervised clustering and visualization with uniform manifold approximation and projection (UMAP) identified 13 distinct cell populations (Fig. 4a, c and Supplementary Fig. 4b).

We constructed a combined single nuclei RNA-seq (snRNA-seq) data sets from adult and fetal human brains32–34, and performed integration and label transfer with our data (Fig. 4b and Supplementary Fig. 4b–e). Based on the label prediction and cluster assignment probabilities calculated using the integrated data set, we predicted seven sub-clusters of OPCs, two of OLs, two of astrocytes, and additional clusters for neural precursor cells (NPC) and pericytes in our data set (Fig. 4a, e, f and Supplementary Fig. 4e). Although the three sub-clusters at the bottom of the UMAP expressed OPC markers, they had enriched expression of cell-cycle-related genes (TOP2A, PCNA, MKI67, and MCM6) and they were in the G2/M/S phase of the cell cycle (Fig. 4d and Supplementary Figs. 4h, 5a); therefore, they were labeled as cycling progenitors35. More mature cells branch out to form more discrete clusters and seem to represent astrocytes and oligodendrocytes that are in the G1 phase of the cell cycle (Fig. 4a, d). Mature OL markers such as MAG, MOG, and ZNF488 were more enriched in cluster OL2 (Fig. 4g, Table 1 and Supplementary Fig. 4h), suggesting that it represents a more mature OL population than cluster OL1. Notably, the OL population in our hESC-derived cells share similarity with all 6 of the OL sub-populations (Fig. 4b, e and Supplementary Fig. 4e–g)32. Gene set enrichment analysis (GSEA) showed enrichment of lipid biosynthesis, enshentment of neurons, and synaptic signaling in both OL populations whereas extracellular matrix, biological adhesion, and protein homooligomerization-related genes were enriched in the AS populations (Fig. 5a, b and Supplementary Data 3).

A relatively isolated, mature cluster at the top of the UMAP contained AS and pericyte sub-clusters (Fig. 4a). The smaller sub-cluster expressed a number of pericyte marker genes (COL1A1, COL1A2, COL3A1, ATAC2)36,37 that are responsible for collagen and extracellular matrix formation and organization, organ morphogenesis, and cardiovascular development (Fig. 4g, Table 1 and Supplementary Fig. 5c, Supplementary Data 3). This group of cells likely represent the VLMC-pericyte cells that are reported to arise from PDGFRα+ precursors38. The other sub-cluster shared similarity to a subgroup of mature astrocytes (Fig. 4e, f) marked by the enriched expression of GFAP, EFN3, and TFF3 (Fig. 4g and Supplementary Data 1).

scRNA-seq provides insight into the PDGFRα-reporter OPC sub-populations. Differential gene expression and GSEA within the OPC sub-clusters revealed a number of differences between them (Fig. 5c, d and Supplementary Data 2, 3). OPC1 shared similarity to a previously reported oligo 4 sub-cluster32 (Fig. 4e), but it also expressed a number of genes (STMN2, NREP, MAP1B, SOX11) involved in neurogenesis, and neuron differentiation (Table 1), which suggests that the cells in this OPC sub-population may have the capacity to differentiate into neurons. OPC3 was enriched for cytokine-mediated signaling pathways, immune response, and defense response genes (IFIH6, ISG15, IFIT1, HLA-A, B and C), and likely represents a sub-population of OPCs that can respond to cytokines39–43 (Fig. 5d–f and Supplementary Fig. 5d, e). OPC4 showed enriched expression of gliogenesis (SOX6, SIRT2, SOX10) as well as neurogenesis (HEY1, TRO, HIS1)-related genes. Receptor protein tyrosine kinase signaling pathway, amide biosynthetic process, and peptide biosynthetic process-related gene sets were also enriched (HES1, DDIT4, FGF12, ZFP36L2, RPL39) in this cell population (Fig. 5d and Table 1). The OPC5 sub-cluster had enriched expression of a number of OL markers and biological functions such as lipid biosynthesis and enshentment of neurons that are similar to that seen with the OL clusters. The cells in this sub-cluster were also in S phase and are enriched for cell cycle and cell division gene sets (Fig. 4d and Table 1), which indicates that they might be a unique group of proliferative cells at an intermediate oligodendrocyte state. The differential gene expression analysis showed that OPC3 and OPC4 shared a number of enriched genes with astrocytes and OLs, respectively, which indicates that these sub-populations of OPCs might already be inclined towards either astrocyte or OL lineages (Fig. 5c).

Since the CyP1 and CyP2 expressed OPC marker genes and also showed similarity to the OPCs by cluster assignment probability (Fig. 4e, f), they were included in the analysis as OPC sub-clusters. However, mitotic processes and cell cycle events were dominant in these groups of cells (Supplementary Fig. 5b). OPC2, on the other hand, expressed OPC markers, and although it expressed only a limited number of enriched genes, a large set of genes were downregulated compared to the other OPCs (Fig. 5d and Table 1).
**Fig. 3 Maturation of reporter OPCs into myelinating oligodendrocytes.**

**a** Image taken with ×40 objective shows PDGFRα localized to the cell membrane and tdTomato in the cytoplasm. Images for individual channels are included in Supplementary Fig. S3c. Scale bar: 100 μm. Immunohistochemistry with PDGFRα was independently repeated three times with similar results.

**b** Live-cell immunofluorescence staining shows that the differentiating reporter OPCs (red) at day 85 express O4 antigen (green). Scale bar: 200 μm. This was independently repeated three times with similar results.

**c** Flow-based quantification of the imaged cells shows ~18% O4+ cells, of which ~14% are also tdTomato+. d Flow analysis of tdTomato expression of day 115 before (left) and after (right) MACS-based immunopurification using O4-antibody-conjugated microbeads. After purification, the population was enriched to 87% tdTomato+ cells. PE is X axis represents PDGFRα-tdTomato. The same gate (left panel of c) was used for both c and d flow analysis. Gating strategy is further detailed in Supplementary Fig. S1j and described in the “Methods” section.

**e** Day 80 Thy1.2 purified population of tdTomato enriched cells were further cultured on PLO-laminin-coated plates (e) or on electrospun nanofibers of 675 nm diameter (f). Scale bar: 400 μm, f left panel, scale bar: 400 μm. Cultures were grown in mitogen-free media for 3 weeks and immunostained with MBP antibody. e Independently repeated five times and f was independently repeated twice with similar results.
Potential of PDGFRα+ precursor cells to differentiate into cells expressing astrocyte markers. Of all the single cells analyzed, 901 (27.5%) expressed GFAP and 733 (22.4%) expressed MBP, markers for astrocyte lineage cells (ALCs) and OLLCs, respectively (Supplementary Fig. 5f). The number of OPCs decrease while the number of OL and AS cells increase over time as they mature (Supplementary Fig. 5f–h), which supports a previous report that PDGFRα+ human OPCs have the potential to mature into both astrocytes and oligodendrocytes in vitro11. We also analyzed the tdTomato+ cells from days 77, 89, and 104 separately (Supplementary Fig. 6a–c). Cells from each timepoint formed distinct clusters with OPC, OL, and AS enriched populations. Although
Validation of findings, particularly related to the astrogenic potential of OPCs and the sub-populations of OPCs. To confirm the reproducibility of our single-cell transcriptomic and other findings on the astrogenic potential of hOPCs and on the sub-populations of hOPCs, we further analyzed the single-cell transcriptomes of additional stem cell-derived hOPCs. This validation analysis included: (1) an independent non-reporter hiPSC line differentiated into OPCs (iP-OPCs) and sampled by a different laboratory group, and (2) an independent batch of hOLLCs differentiated from the PD-TT reporter line (D85-OLLCs) (Fig. 6a and Supplementary Figs. 7–9). Sixty-day-old iP-OPCs were FACs purified using PDGFRα antibody (Supplementary Fig. 7a) and day 85 PD-TT reporter OLLCs were MACS purified with thyl1.2 microbeads prior to single-cell capture using 10x platform. Although at the D60 timepoint, as expected, there were only a few cells that expressed mature OL and astrocyte markers, the co-embedding of the hiPSC-OPC data set with the OLLC data set from Fig. 4 indicated that the iP-OPCs were fully consistent with the trajectories we defined with the PD-TT-derived reporter OLLCs (Fig. 6a and Supplementary Fig. 7b).

Additionally, expression of cell-cycle-related genes (e.g., PCNA, TOP2A, MKI67), OL-related genes (CNP and TCF7L2), as well as TFs that are known to drive astrocyte differentiation (SOX9, NFIA, and NR2F1), were present in these D60 cells (Fig. 6b and Supplementary Fig. 7c). GSEA analysis on the iP-OPC data set also showed cluster-specific enrichment of OL and AS-related pathways and highlighted sub-populations of OPCs that are cytokine-responsive and OPCs that are inclined to either OL or AS lineages (Fig. 6c and Supplementary Fig. 8, Supplementary Data 7). Additionally, the D85-OLLCs integrated well with the reference data set from Supplementary Fig. 4c, d (Fig. 6c–f). The OL, AS, pericyte, cycling progenitor cell populations, and different OPC sub-populations including the cytokine-responsive OPC sub-populations, were present in the D85 data set as well (Fig. 6e–h and Supplementary Fig. 9a–e, Supplementary Data 6).

In addition, we also performed scRNA-seq analysis of day 89 PD-TT reporter OPCs purified by an independent method, using O4-antibody-conjugated microbeads. Similar to the PDGFRα expressing cells, a sub-population of the O4+ cells also express astrocyte markers (Supplementary Fig. 6d).

**Pathway enrichment analysis reveals pathways associated with OPC sub-populations, and oligodendrocyte and astrocyte differentiation.** In order to get further insight into the molecular pathways that define OPCs, OLLCs, or ALCs sub-populations, we performed pathway enrichment analyses on differentially expressed genes from each cluster using Ingenuity Pathway Analysis (IPA). In the mature OL2 cluster, a number of pathways including CXCR4, Sphingosine-1-phosphate (SIP), and integrin signaling pathways (ISP), which are known to be important for OPC maturation, oligodendrocyte survival, and myelination, were upregulated and, ILF2, ILK, and Estrogen receptor signaling were downregulated, with strong z-scores (Fig. 7d). Numerous pathways with P-values denoting significance were also identified for the OPC sub-clusters, and OL and AS clusters (Fig. 7a–c and Supplementary Fig. 10a, b, Supplementary Data 4). The cholesterol biosynthesis pathways (CBPs) and the cholesterol biosynthesis intermediate (CBI) such as mevalonate and zymosterol signaling, known to be involved in OL differentiation and myelination, were enriched in both the OL clusters. The mTOR, eIF4 and P70S6K pathways were differentially downregulated between the OL1 and more mature OL2 sub-clusters (Fig. 7a). The Cdc42 and Rho-family GTPases signaling, Phagosome maturation and Caveolar-mediated endocytosis signaling, and Ketolysis, which are implicated in astrocyte functions10–14, were enriched while the CBPs, Epoxyquinoline, and Wnt/b-catenin genes are highly enriched and differentially expressed genes from the OLLCs and ALCs populations are consistent with the previous reports. Genes that have been reported to be specifically enriched in human OLs and astrocytes but not expressed in mouse OLs or astrocytes, such as APCDD1, HMGCS1, PMP2, and WIF1, were also enriched in the respective clusters as expected. We assessed the overlap of genes enriched in our hESC-derived OLLC, ALC, and OPCs to that of mouse OLLCs, ALCs, and OPC-enriched genes. Although numerous genes including known marker genes overlapped between the two data sets, we noted several differences (Fig. 5g, Supplementary Fig. 9f, Source Data).

We additionally identified a number of differentially expressed primary microRNAs (pri-miRNAs) in our scRNA-seq data (Supplementary Fig. 6f). For example, miR219-A2 was enriched in the OL clusters while miR100HG and miR99AHG were enriched in the astrocyte clusters. These differentially expressed pri-miRNAs are potentially involved in regulating the fate of an OPC to become either an OL or an astrocyte.
| Cycle 1 | Cycle 2 | Cycle 3 | Cycle 4 | Cycle 5 | Oligodendrocytes |
|--------|--------|--------|--------|--------|------------------|
| OPC1   | OPC2   | OPC3   | OPC4   | OPC5   | Peri             |
| NUF2   | CDK1   | SMC2   | SOX11  | NKAIN4 | HS3ST1          |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| ASPM   | HSP90A1| NRP1   | CCNB1  | ZFP36L2| RP11-89N17       |
| CENPF  | FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |
| HIST1H4C| FABP7  | CAMR1  | NLZ    | PRDX1  | BCAS1           |

The entire list of genes can be found in Supplementary Data 1.

OPCs oligodendrocyte progenitor cells, At astrocytes. Peri-pericytes.
signaling pathways were comparatively downregulated in the astrocyte cluster (Fig. 7b). EIF2 signaling demonstrated a wave-like pattern. It was downregulated in the early progenitors, upregulated in OPCs, and downregulated again in the OLs and AS populations (Fig. 7a–c and Supplementary Fig. 10a, b).

We also performed IPA on the OPC sub-clusters. In one analysis we looked at the genes differentially expressed between the OPC sub-clusters vs all the other clusters. In another comparison, we focused on the genes differentially expressed between the individual OPC sub-clusters. Both analyses (Fig. 7c–f)
and Supplementary Fig. 10b) showed enrichment of antigen presentation and interferon signaling in OPC3. The Cdc42 and Caveolar-mediated endocytosis signaling and phagosome matura-
tion were also upregulated in this sub-cluster, which further supports that the cells in this sub-cluster are inclined towards astrocyte lineage. EIF2, BEX2, mTOR, Notch, Sirtipuin, and Wnt signaling were upregulated and the antigen presentation and interferon signaling were downregulated in OPC4. A number of cholesterol and GGPP biosynthesis-related pathways, that are implicated in OL maturation, were enriched in the OPC5. Nitric oxide and iNOS signaling, and nNOS signaling, were comparably upregulated in OPC3 and OPC5. We also found upregulation of the long non-coding RNA (IncRNA)-HOTAIR pathway, which is known to modulate PI3K/AKT/mTOR signaling, as well as enrichment of the mTOR signaling in the OPC1 sub-cluster. As expected, the cycling progenitors were enriched for the cell-cycling-related pathways and NER, BER, and ATM signaling, which were downregulated in all but OPC5 (Supplementary Fig. 10a). The pathway enrichment analysis further supports the result from the GSEA that the OPC sub-
populations are functionally diverse and developmentally heterogeneous.

Of potential interest, we observed enrichment of genes related to coronavirus pathogenesis pathway (CCND1, EEF1A1, RPS29, FOS, STAT1, E2F3) in the OPC3-5 (most significantly in OPC4) sub-clusters (Supplementary Fig. 10b). This finding suggests the possibility that the SARS-CoV-2 virus, which is causing the current COVID-19 pandemic, may have the capacity to infect CNS progenitor cells. This possibility is consistent with the recent report that iPSC-derived BrainSpheres can be infected with SARS-CoV-2, and that the infected BrainSpheres can support viral replication.

To confirm the biological relevance of a number of above-described, bioinformatically implicated pathways in hOPC to hOL maturation, we experimentally tested the consequences of pharmacologically inhibiting their activity in developing hOPCs. Similar to a recent report studying murine OL differentiation, inhibition of CYP51A1, TM7SF2, and EBP, which are associated with the CBP, increased expression of MBP transcripts in our human system (Fig. 7c). Immunostaining confirmed stronger MBP expression and more MBP+ cells in these samples (Fig. 7f). Inhibition of mTOR signaling by rapamycin reduced the expression of MBP and increased the expression of the OPC markers PDGFRα and CSPG4 (NG2), which indicates that the mTOR signaling, although downregulated in the OLs, is still essential for the maturation of hOL from hOPCs. However, targeting the S1P and CXCR4 pathways, at least with the compounds we tested, did not show any significant effect on hOL differentiation/maturation.

Pseudotemporal trajectory analysis further defines the bipotential nature of PDGFRα+ cells. We performed monocle-based pseudotime analysis on our scRNA-seq data to create a develop-
mental trajectory tracing the lineage specification of PDGFRα+ cells as they mature. Analysis of the pseudotemporal trajectory presented two prominent paths for the precursor cells, suggesting that the PDGFRα+ cells can follow two distinct cell lineages (Fig. 8a). We examined the highly differentially expressed genes and transcription factors (TFs) between the two trajectories and identified path II as OLLCs and path III as ALCs using the Branch Expression Analysis Modeling (BEAM) regression model (Fig. 8b, c).

The monocle-based developmental trajectory analysis defined seven distinct cellular states (Fig. 9a). Cells in state 1 were precursor cells, state 2 was composed of astrocyte cells, and state 6 consisted of oligodendrocyte cells. Analysis of the pseudotemporal expression pattern of the OL and astrocyte genes indicates that the OL genes MBP and PLP1 have similar kinetic trends and both are enriched in state 6 cells, while the astrocyte markers GFAP and AQP4 are enriched in state 2 cells (Fig. 9c). State 5 and 7 cells diverged from the OL trajectory, and were interestingly enriched for astrocyte markers rather than OL markers, and share a stronger correlation with the state 2 astrocyte cells (Fig. 9b and Supplementary Fig. 11a). We also performed pseudotime analysis on the day 89 O4+ cells. Similar to the purified PDGFRα-
tdTomato+ OPCs, the O4+ cells also showed two potential lineages as indicated by OL vs astrocyte trajectories and the kinetics of gene expression within the trajectories (Supplementary Fig. 11d-f).

We also examined the most differentially expressed genes and transcription factors at nodes 1, 2 and 3, and in cells at states 2 vs state 6, using the BEAM regression model (Fig. 9d, e and Supplementary Fig. 12a, b). This analysis identified numerous differentially expressed genes of potential interest. In addition to TFs previously implicated in OPC/OL/astrocyte differentiation, such as SOX10 and TCF7L2, this analysis implicated a number of previously less characterized factors in OPC/OL/AS differentiation, including ZEB2, TSC22D4, ARID4B, PARP1, E2F3, and ARHGA35, which were significantly enriched in OLs, and HES1, FO8B, NFIA, NR2F1, and ZFP36L1, which were enriched in astrocytes. Numerous TFs, such as SOX4, SOX11, MLLT11, RBM22, ZNF711, EZH2, and DACH2, were slightly upregulated in OLs but highly downregulated in ALCs (Figs. 8b, c and 9d, e).

A number of these genes are already enriched in day 60 hiPSC-
derived OPCs (Fig. 6b, Supplementary Fig. 7d). Furthermore, each path of the trajectory consisted of cells from all three (days 77, 89, and 104) timepoints (Supplementary Fig. 12c), which further supports that PDGFRα+ cells are transcriptionally and developmentally heterogeneous, with some cells, even at the earliest timepoint tested (day 77), being already relatively mature.

Discussion

Recent advances in stem cell biology and differentiation metho-
dology have led to the development of protocols for the genera-
tion of hOLCs from hPSCs, and hold the promise of aiding in the development of remyelination-based approaches for the treatment of MS and other demyelinating diseases. The ability to promote remyelination is also highly relevant for ongoing work in regenerative medicine, such as efforts to promote optic nerve regeneration. In current human stem cell differentiation
protocols, hOPCs generally make up <50% of the resulting differentiated cell population\textsuperscript{12,20,21}, often requiring FACS-based purification for downstream studies. Here, in order to develop a simple and efficient method for obtaining highly purified populations of PDGFR\(\alpha\)-expressing hOLLCs, we used CRISPR-based genome editing to introduce an IAP (P2A-tdTomato-P2A-thy1.2) tag into the endogenous PDGFR\(\alpha\) locus. The resulting reporter cell line allows for optimization, scalable differentiation, and purification of hOLLCs (>90% PDGFR\(\alpha\)/tdTomato-expressing cells) at different stages of differentiation. The cells purified by this method maintain high survival, competence, and the capacity to mature into MBP\(^{+}\) OLs even after long-term cryopreservation.
This reporter system provides a useful resource and a powerful tool for optimizing new, more efficient differentiation protocols, enabling the easy quantification of effects that small molecules and patterning factors have on promoting OPC differentiation. In addition, the high capacity and simplicity of the hOPC purification method could aid in establishing a human OPC-based drug-discovery platform for performing high-throughput screens for myelination promoting compounds.

PDGFαα expressing OLLCs from the mouse CNS have been well characterized using scRNA-seq analysis, but similar analysis using human cells has not been reported. To define the transcriptional diversity and heterogeneity of PDGFαα expressing human OLLCs, we performed scRNA-seq on hESC-derived reporter cells purified at various timepoints after the initiation of differentiation. We also analyzed PDGFαα expressing cells from an independent, non-reporter hiPSC line. Unbiased clustering of the cells identified distinct clusters of OPCs, astrocytes, oligodendrocytes, and pericytes. We also identified sub-populations of OPCs that were defined by their developmental stage and commitment to either the OL or astrocyte lineages. A sub-population of hOPCs with enriched cytokine response signaling was also uncovered.

A variety of studies indicate that OPCs and pre-OLs, but not mature OLs, express PDGFαα. Thus, it was surprising that a significant number of our PDGFαα+tdTomato+ cells showed a strong correlation with OLs and astrocytes and expressed numerous mature markers (Fig. 4 and Table 1). It should be noted, however, that the PDGFαα mRNA in our single-cell data are not well detected in the clusters representing mature cells (Fig. 4d). Possible explanations, given that the tdTomato and Thy1.2 proteins were obviously present at the time of purification, are that (1) PDGFαα mRNA is not actively expressed in the mature cells at the time of purification, (2) the half-life of tdTomato is longer than PDGFαα, (3) allowing purification of cells that no longer express PDGFαα mRNA, or (4) perhaps, in the mature cells PDGFαα mRNA is expressed as a low abundance transcript, and such transcripts are often missed in scRNA-seq.

Although the Dropseq approach we used does not capture mature miRNAs because they are not poly-adénylated, Dropseq can sometimes capture pri-miRNAs. We identified a number of pre-miRNAs including pre-miR219-2A, which is highly enriched in human OLs and has previously been shown to be important for myelination and remyelination in mice. Our PT-DT reporter system could thus be a useful resource for future studies to more directly explore ALC and OLLC specific miRNAs.

Additionally, the utility of our scRNA-seq data is further supported by IPA analysis, which revealed pathways associated with hOL and astrocyte differentiation. Of particular interest is the finding of enrichment of the CXCR4, Sphingosine-1-phosphate, integrin, mTOR, and cholesterol biosynthesis signaling pathways in mature OL cells. Pharmacologically modulating the cholesterol biosynthesis pathway (CBP) increased the amount of MBP mRNA and enhanced OL differentiation. Inhibition of CBP has been shown to cause accumulation of 8,9-unsaturated sterols. Since these compounds increase the amount of MBP mRNA, it is possible that the 8,9-unsaturated sterols target upstream of MBP to increase its production. Transcriptomic study of the human OPCs treated with these compounds or supplemented with the 8,9-unsaturated sterols would help identify the upstream regulators of MBP expression. Moreover, further studies to modulate the other pathways identified in this study could potentially help in remyelination-based drug-discovery efforts.

The long-term expression of PDGFαα-ttdTomato/Thy1.2 in the reporter OLLCs allowed us to perform pseudotime analysis and study their differentiation trajectories. The pseudotime analysis revealed OLs vs astrocytes as the two major lineage trajectories of the PDGFαα+ hOPCs. Oligodendrogenesis and astrogenesis from our PDGFαα-ttdTomato+ cells are similar to in vitro cultured, PDGFαα+ primary hOPCs, but contrasts with a recent report that suggests that Pdgfra+ mouse OPCs can give rise to OLs, neurons or VLMC-pericytes but not astrocytes in vivo. We did not find any study that discusses if astrocytes can arise from in vitro cultured Pdgfra+ mouse OPCs. However, data suggesting that Sox10+ mouse OPCs can differentiate into astrocytes in vivo as well as in vitro have been presented. Our analysis, although keeping in mind that it is an in vitro study with hESC-derived PDGFαα+ and O4+ cells, and that the findings need to be confirmed in vivo, also suggests that astrocytes can originate from hOPCs.

We also identified transcription factors (TFs) that potentially help modulate lineage specification, differentiation, and maturation of OPCs into either OLs or astrocytes. Continued upregulation of the TFs ZEB2, TSC22D4, ARID4B, PARP1, E2F3, SOX10, TCF7L2, TSC22D1, RBP1, ARHGA35, SOX4, SOX11, MLLT11, RBM22, ZNF1711, EZH2, and DACH2, which are enriched in OLs and downregulated in astrocyte cells, and HES1, EGR1, FOSB, NFIA, NR2F1, ID3, KLF6, and ZFP36L1, whose expression is enriched in astrocytes and decreased in OLs, seems to drive specification of OLs vs astrocytes from PDGFαα+ OPCs. The role of the majority of these TFs in OL/astrocyte differentiation and maturation has not been studied. Since the function of SOX10 and TCF7L2 in OL development and NFIA and NR2F1 in astrocyte differentiation is well known, and a crucial role of ZFP36L1 in OL-astrocyte lineage transition was recently reported, it is conceivable that the TFs we identified have important roles in OL vs astrocyte lineage specification. Loss of function and gain of function studies of these genes and TFs in OPCs will help to further validate their potential role in lineage specification of...
human OPCs and provide valuable information to our current understanding of OL and astrocyte differentiation.

Although mouse and human OPCs and OLs share transcriptomic similarity and conserved pathways, there appear to be some important species-specific distinctions with respect to their transcriptome, development, and regeneration. In vivo, human gliogenesis starts during gestation and myelination can occur until age 22; while in rodents, gliogenesis occurs postnatally and myelination is resolved by day 60 [69,70]. In vitro, Ascl1 and FGF2 promote the generation of OLs from rodent OPCs but not from human OPCs [10]. In addition, unlike in animal models of MS where myelin is regenerated by newly formed oligodendrocytes, the capacity to generate oligodendrocytes around the lesions of human patients is generally diminished, and the limited remyelination that does occur at MS lesions is likely generated by pre-existing OLs [22]. Therefore, for improved disease modeling and to...
support drug-discovery efforts, it is crucial to expand our understanding of human OL maturation and better define the similarities and differences between murine and human OL differentiation. We hope that the hOPC purification system described in this manuscript along with our scRNA-seq data set (http://zacklab.org/OPCs/) will help provide the basis for ongoing and future studies that will more fully define the molecular mechanisms of human OL differentiation, maturation, and myelination.

Methods

Human pluripotent stem cells (PSCs) and culture conditions. hESC line WA09 (WiCell), an NIH-approved hESC line (NIH approval number: NIHhESC-10-0062 and WiCell agreement number: 09-W571), was used for this study. hESCs were maintained in StemFlex media (A3349401, ThermoFisher Scientific) on growth factor-reduced Matrigel (354230, Corning) coated plates at 37 °C, 10% CO2/5% O2. However, during reporter cell-line generation, the hESCs were maintained in mTeSR1 media (Stemcell Technologies). The iPSCs were always maintained in mTeSR1 media in normal culture conditions (37 °C, 5% CO2). hPSC colonies were passaged by dissociating with Accutase (A6964, Sigma-Aldrich). Cells were maintained in stem cell media containing 5 mMrapamycin (WZ811), or mTOR (rapamycin) signaling pathways. The treated cells were tested by qPCR to assess the effect on hOPC differentiation to OL. Two technical and two biological replicates were used for each qPCR. Biological and technical replicates are distinguished by solid vs open symbols. Data are presented as mean ± SD. Source data are provided as a Source Data file. Immunostaining of the cells treated with different compounds shows stronger MBP staining and more MBP+ cells in the samples treated with CBP inhibitors. Scale bar: 275 μm. Immunostaining was independently repeated twice with similar results.
blebbistatin (B0560, Sigma-Aldrich) for the first 24 h after passaging, to improve single-cell survival.

Karyotype analysis was performed using a qPCR based hPSC Genetic Analysis Kit (StemCell Technologies, #07550). If significant aberration was detected, further analysis with G-banding (CellLine Genetics) was performed. Chromosomal aberration in hPSCs is common, with hESCs often having duplication in chromosome 1q and 20q as they get to higher passage number. We also observed isochromosome duplication of chromosome 1q as early as passage 10, ~3 months after generation of the line. Of note, the WA09 cells were already at p30 when purchased from WiCell. 10 out of 20 clones analyzed by G-banding had this abnormality. Cells were routinely tested for mycoplasma contamination (MycoAlert, Lonza) and only the cells free of contamination were used for OPC differentiation.
Cloning. A guide sequence targeting the stop codon of the PDGFRα locus was designed in Deskgem.com. Guide sequence with minimal off-target and very high activity score was chosen and cloned into the BbsI restriction site of Addgene #62988). *pScas9(BB)-2A-Puro (PX-459) V.2.0 was a gift from Feng Zhang (Addgene plasmid #62898; http://n2t.net/addgene/62898). To clone the donor plasmid, a ~2 kb PCR product was amplified from genomic DNA extracted from H9 ES cells and cloned into Zero Blunt TOPO cloning vector (ThermoFisher Scientific). The tdTomato-P2A-Thy1.2 reporter DNA sequence was then introduced into the TOPO-based donor plasmid, precisely upstream of the PDGFRα stop codon using Gibson assembly (New England Biolabs).

Generation of PDGFRα reporter cell line. Gene editing and reporter cell line generation was performed using a transient antibiotic selection method.27 Cells were transfected using the Lipofectamine Stem (STEM00001, ThermoFisher Scientific) transfection reagent following the manufacturer’s recommended protocol. 0.35 µg Cas9 plasmid (Cas9-P2A-Puro modified from Addgene #62988) containing a gRNA sequence targeting the stop codon of the PDGFRα locus and 0.75 µg of donor plasmid were used for transfection. Roughly 40 h after transfection, the cells were selected with 0.6 µg/mL of puromycin for 24 h. Selected cells were passaged at 500–1000 single cells per well of a 6-well plate for colony formation followed by colony picking and PCR analysis.27, PCR was performed using the Phusion Flash mix mastermix (ThermoFisher Scientific) and a 2-step PCR protocol following the manufacturer’s instruction. To generate the PDFF54 gene, we picked 30 single-cell-derived colonies, of which, 7 had clear homozygous knock-in (Supplementary Fig. S1a). 5 of the homozygous colonies were further passaged and single-cell colonies picked again to confirm the homozygous population of the knock-in cells (Supplementary Fig. S1b). Two clones from each of the 5 lines were then differentiated for 12 days to confirm tdTomato expression under a fluorescent microscope. Since all the clones seem to perform equally well, three clones were further expanded and stored as PDFF54 reporter cell lines. Of these, only one clone was used for differentiation and single-cell transcriptomic analysis.

For off-target analysis, the top 5 potential off-target locations were chosen based on Deskgem.com, the website used to design the guide sequence. Specific primers upstream and downstream of the location were designed (Supplementary Table 1). PCR followed by Sanger sequencing was performed to confirm no off-target mutation was caused during genome editing.

Oligodendrocyte differentiation protocol. hESCs and hiPSCs were differentiated into Oligodendrocyte progenitors (OPCs) using a previously described30 with minor modifications. Briefly, hESCs were dissociated to single cells and plated on Matrigel-coated plate at 100,000 cells/well of a 6-well plate and maintained in StemFlex media at 37 °C, 10% CO2/5% O2. hiPS cells were maintained in mTeSR1 media at 37 °C, 5% CO2. Two days after plating, neural differentiation and spinal cord patterning were induced through dual SMAD inhibition and the addition of 100 nM all-trans RA.31 From days 8 to 12, differentiating cells were maintained in neural induction media supplemented with RA (100 nM) and SAG (1 µM). On day 12, adherent cells were lifted and cultured in low-attachment plates to favor sphere aggregation. On day 30, spheres were plated into poly-L-ornithine/laminin-coated dishes in a media supplemented with B27 (ThermoFisher, 12507010), N2 supplement (ThermoFisher, 1279502-1 R&D systems), neurotrophin-3, HGF (294-HG-025 R&D systems), and T3. Detailed information on each reagent used is listed in supplementary Table 2. Once a significant number of tdTomato+ OPCs were visible around days 65–70, a mitogen-free glial media was used to drive oligodendrocyte maturation.

Cryopreservation and revival of hPSCs and purified OPCs. For cryopreservation, hPSCs were dissociated with accutase, diluted in DMEM/F12, and centrifuged at 150 × g for 5 min. Cell pellets were resuspended at 1 million cells/mL in CryoStor CS10 (07930, StemCell Technologies), placed in cryovials, and stored at −80 °C. MACS purified hOPCs were pelleted by centrifuging at 250 × g for 5 min and cryopreserved in CryoStor at 5 million cells/mL. hPSCs were revived by incubating cryopreserved cell pellets on ice for 3 min, washing two times with RA and resuspending in mTeSR or StemFlex medium supplemented with bbeblatin, and plating them in Matrigel-coated plate. For cryopreserved purified hOPCs, cells were thawed at 37 °C for 3 min, resuspended in of PDGF media, and directly plated on PLO-laminin-coated dish without centrifuging/washing the cells. The cells were then allowed to settle and attach for 2 h. After 2 h, media was changed very gently to remove DMSO, which is part of CryoStor buffer. We noticed that washing the cryopreserved cells can lead to ~40% loss of cells, which fail to precipitate and form pellets. With the direct plating method described here, the purified and cryopreserved OPCs can be revived with >80% retention and survival. However, we noticed a significant loss of cells if passed after purification or after revival from cryopreservation, therefore, we do not recommend passing the hOPCs after purification or revival.

Flow cytometry and MACS purification of the reporter hOPCs. Flow cytometry analysis and MACS purification were performed as previously described with the following modifications. Cells were dissociated into a single-cell suspension by incubating in accutase for ~45 min. The single-cell suspension was then passed through a ~70 µm cell strainer (BD Biosciences), washed, and resuspended in Live Cell Imaging Solution (ThermoFisher Scientific) for analysis with an Echo 550 acoustic liquid handler (Labcyte, CA). For flow analysis, BSC and FSC were used to select and subset live cells, and only live cells were used to quantify a number of tdTomato+ cells. A gate was set up using WT hES cells differentiated to day 95. Since it is not practical to differentiate a WT line every time a reporter line is differentiated, the same gate was used for every flow analysis. For MACS purification, cells were resuspended in MACS buffer after passing through the cell strainer. A CD90.2 (THY1.2), O4, or A2B5 MicroBeads were added to the cell suspension and incubated at room temperature for 15 min for cell binding. Cells were generally run through the MS column twice without additional supplementation of MicroBeads to increase the purity and achieve ~90% tdTomato+ cells. All MACS reagents were purchased from Milltenyi Biotec (Auburn, CA) and manufacturer instructions were followed.

Testing pharmacological compounds on hOPCs. Day 85 hOPCs were MACS purified to ~90% purity. 200,000 purified hOPCs were plated per well of a PLO/laminin-coated 24 well tissue culture plate, in a mitogen-free glial media. Day 85 hOPCs were plated in culture, media was replaced with media containing different compounds or DMSO. Ketokonazole, amolfoline, and tasin-1, which target CYP51A1, TM7SF2, and EBP, respectively, were used to target the cholesterol biosynthesis pathways; CYM5520 was used to target SIP; WZ811 was used to inhibit CCR4 signaling, and rapamycin was used to inhibit mTOR signaling. Based on previous publications and our lab’s unpublished data, the following compounds were used to inhibit CXCR4 signaling: (a) CNM1, (b) GM1 (10 µM) (2) (c) Amorolfine (370 nM) (d) Amorolfine (370 nM) (e) Tamin-1 (41 nM) (f) Rapamycin (123 nM). All the compounds were purchased from Selleckchem (selleckchem.com). Media containing the compounds was replaced on day 3, and on day 7 the cells were either lysed for RNA extraction and qRT-PCR or fixed with 4% paraformaldehyde for immunostaining.

Immunofluorescence staining, microscopy, and qRT-PCR. All sequences for qRT-PCR primers can be found in Supplementary Table S1. Total RNA was isolated using the RNeasy Mini Kit (QIAGEN) and reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). A 25 µL PCR reaction was set up using an acoustic liquid handler (ECCHO 550, Labytec) and performed with the CFX384 real-time PCR instrument (Bio-Rad). Assays included at least two technical and two biological replicates and were run using the Sso Advanced Universal SYBR Green Supernix (Bio-Rad). Primers used for qRT-PCR are listed in Supplementary Table S1.

For immunofluorescence staining, cells were fixed with 4% paraformaldehyde, and simultaneously permeabilized and blocked with 0.2% Triton X-100 + 5% BSA + 5% normal goat serum (or serum specific to the host of secondary antibody) for an hour. Cells were then incubated with the appropriate dilution of primary antibodies overnight followed by secondary antibodies for 2 h. Antibodies used for the study are listed in Supplementary Table 2. Fluorescence images were taken using either the EVOS FL Auto 2 (ThermoFisher Scientific) or Zeiss 510 confocal microscope. Co-staining analysis of the tdTomato and PDGFRα was performed...
using a built-in algorithm of the ArrayScan image analysis software (ThermoFisher Scientific ArrayScan XTI). Thirteen images were used for analysis.

Live cell imaging. The EVOS FL Auto 2 Cell Imaging System (ThermoFisher Scientific) was used for imaging cells in culture over time for time-lapse videos. Cells were maintained in a live cell chamber at 37 °C with 5% CO2 and 85% humidity and areas around nephrospheres were scanned every 20 min. Images were collected at 15 fps using Image) to generate time-lapse videos.

**Statistical analysis.** All qRT-PCR data are presented as fold change in RNA normalized to the expression of two housekeeping genes: either GAPDH and SRT72 or GAPDH and CRRB2. qRT-PCR data were analyzed using CFX Maestro (Bio-Rad) qPCR analysis software and graphed using Prism (GraphPad, V9).

Drop-seq-based single-cell capture and RNA-sequencing. Drop-seq-based single-cell RNA-seq was performed as previously described by Macosko et al.20. Barcoded microparticles were purchased from Chemgenes Corporations. During differentiation, a large number of differentiating cultures are PDGFRA+ on day 75, providing sufficient cell numbers for downstream analysis.42,43 We observe a similar phenomenon with tdTomato-expressing cells in our differentiation culture. Therefore, day 75 and two other timepoints, two weeks apart were considered for time-course single-cell capture. Since MACS purification and Drop-seq could not be timed on the exact intended day, we ended up using cells from days 77, 89, and 104. The cells were washed and resuspended in a single batch withison and O4+ cells are from two different batches of differentiation. Small fraction of the purified cells was used for FACs-based analysis to confirm that >90% of them were tdTomato. Cells were MACS purified 2–3 times until ~90% purity was achieved. Using the microfluidic reporter cells were captured into ~1 nl size droplets containing barcoded nanoparticles and lysis buffer. Generated droplets were broken with perfluorooctanol (Sigma, 370533) in 30 ml of 6× SSC. The beads were then washed, reverse transcribed, PCR amplified, and the amplified cDNA quantified using a Bioanalyzer High Sensitivity Chip (Agilent). The cDNA was then fragmented and amplified for 3′ prime sequencing with the endogenous cDNA sample prep kit (Illumina). cDNA and libraries for each sample were prepared independently, and an equimolar amount of each library was then pooled together for sequencing. The libraries were purified, quantified, and then sequenced on the rapid flow chip in Illumina HiSeq 2500. From the three independent timepoints, a combined total of ~4800 purified cells were captured.

Bioinformatic analysis

**Quality control and clustering and differential gene expression.** The principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE) analyses were performed using a previously published R package, Seurat21. As a quality control, only cells that had a minimum of 250 mRNA molecules and a maximum of 20% mitochondrial RNA were used for analysis. In order to eliminate probable doublets from the data set, we bioinformatically filtered out cells with >30,000 unique molecular identifier (UMIs) (Supplementary Fig. 53a). Additionally, we removed cells that exhibited (1) expression of <250 genes or (2) >20% mitochondrial RNA were used for analysis. 1874 highly variable genes were input for PCA analysis, and the 16 statistically significant PCs were used for clustering and UMAP or t-SNE embedding. To identify differentially expressed genes for each cluster, genes that are expressed in less than 10% in either group or have a log fold change of <0.25 are filtered out. CyP1 was G2M phase while CyP2 and CyP3 are in S phase. To conduct GSEA analysis on the differentially expressed genes, enriched genes are first ranked by the product of −log10(P-value) and log (average fold change). The gene ranks are used for pre-ranked GSEA analysis using all Gene Ontology terms, with 1000 permutations.44

Pseudotemporal trajectory analysis. Time-series analysis to generate a pseudotemporal trajectory was performed using an unsupervised differential gene expression test based on sample age in Monocle, following previously published gene expression test based on sample age in Monocle. Following previously published pseudotemporal trajectory analysis, the top 14 clusters were then given a new cluster label, with 7 Oligodendrocyte clusters, 3 Astrocyte clusters, 2 OPC clusters, 1 Neuro Progenitor cluster and 1 Pericyte/Endothelial cluster being assigned. Next, we carried out integration and label transfer of cells in our data set with the reference using Seurat Anchor (3.0) with CCA dimension reduction using a total of 1568 genes that are variable in the reference data and also detected in our data. The cluster assignment probabilities are then transferred for each cell in our data set by weighting labels of 20 closest neighbors in the CCA space (Supplementary Fig. S4), and for each cluster in our data set, an average of the label assignment probability is calculated (Fig. 4e, f).

To generate a co-embedding of tdTom and the reference (Fig. 4b), normalized gene expression values of the variable genes are also transferred for each cell in the tdTom data similar to the labels. The resulting data are then combined with the reference gene expression. Gene-wise scaling, PCA, and UMAP are subsequently performed on the combined data.

G2M score calculation and gene set enrichment analysis. Cell cycle scoring was performed with a list of cell cycle genes collected from Tirosi et al. (Supplementary Table S3). The cycling progenitor clusters are cycling and the other clusters were non-cycling. CyP1 was G2M phase while CyP2 and CyP3 are in S phase. To conduct GSEA analysis on the differentially expressed genes, enriched genes are first ranked by the product of −log10(P-value) and log (average fold change). The gene ranks are used for pre-ranked GSEA analysis using all Gene Ontology terms, with 1000 permutations.

Calculation of Spearman correlations and human-mouse overlap assessment. Total expression of all the genes expressed in both our data set and the previously published snRNAseq data set (13,658 total genes) was used to normalize gene expression in each cell. The normalized expression was averaged across each population and 1195 variable genes expressed in both data sets were then used to calculate the Spearman correlation between each population. To combine all the oligodendrocyte sub-population as one population, expression was weighted by the number of cells in each sub-population. A similar weighted average was also used to combine the two astrocyte sub-populations. For the comparison with bulk RNAseq data set, 767 highly variable genes from each human tissue-type published by Zhang et al.45 were compared to each of our population’s averaged normalized expression. For comparison with mouse cells, all the genes expressed in each of our clusters were compared to the genes from each of the mouse CNS cell-type previously published2.

To assess the overlap of enriched genes in our hESC-derived OLLCs, ALCs, OPCs, and endothelial/pericyte cells to that of the enriched genes in the corresponding population from the mouse data set, all 1266 differentially expressed genes used for the expression heatmap (Fig. 4b) were used for the comparison. From the mouse data set, genes expressed at levels >20 FPKM and >5-fold relative to average expression in all other populations, were considered enriched. Codes for the bioinformatic analysis of the single-cell data are deposited in GitHub.

**The networks and pathway analyses.** The networks and pathway analyses were generated through the use of Ingenuity Pathway Analysis (IPA QIAGEN Inc)47. For the analysis, differentially expressed genes and their corresponding P-value and fold change (from Supplementary Table 1) were uploaded for each cluster. The steroid biosynthesis pathway map was generated using KEGG via Visualization and Integrated Discovery (DAVID) tools and its steroid biosynthesis pathway as the reference pathway.

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

**Data availability.** RNA-seq data generated for this paper are deposited to the NCBI’s Gene Expression Omnibus (GEO) database with accession number GSE146373. Following publicly available data sets (GSE118257, GSE104276, GSE52564, and GSE73721) and Bioproject (544731) were also used for this study. Figures associated with the RNA-seq data are Figs. 4–8 and Supplementary Figs. S4–S12. List of genes, GSEA, and IPA for each scRNA-seq based cluster are available as Supplementary Data 1–7. GSEA and gene expression data are also available for viewing at https://raklab.org/OPCs/. Source data are provided with this paper.

**Code availability.** The codes used to generate the results reported in this manuscript can be accessed at https://github.com/wefang/TdTom76. Previously unreported custom codes were not used in the manuscript.
References
1. Baumann, N. & Pham-Dinh, D. Biology of oligodendrocyte and myelin in the mammalian central nervous system. *Physiol. Rev.* 81, 871–927 (2001).
2. Dugan, J. D. & Raff, M. A. Inherited and acquired disorders of myelin: the underlying myelin pathology. *Exp. Neurol.* 283, 452–475 (2016).
3. Zhang, Y. et al. An RNA-sequecing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947 (2014).
4. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* 352, 1326–1329 (2016).
5. de Faria, O. J., Mozafari, S., Baron-Van Evercooren, A. & Kuhlmann, T. Targeting human oligodendrocyte progenitors for myelin repair. *Exp. Neurol.* 283, 489–500 (2016).
6. Elbaz, B. & Popko, B. Molecular control of oligodendrocyte development. *Trends Neurosci.* https://doi.org/10.1016/j.tins.2019.01.002 (2019).
7. Dietz, K. C., Polanco, J. J., Pol, S. U. & Sim, F. J. Targeting human oligodendrocyte progenitors for myelin disease. *Nat. Biotechnol.* 34, 1143–1155 (2016).
8. Wang, J. et al. Transcription factor induction of human oligodendrocyte progenitor cells. *Cell Transplant.* 23, 1685–1700 (2014).
9. Sim, F. J., Windrem, M. S. & Goldman, S. A. Fate determination of adult human glial progenitor cells. *Neuron Glia Biol.* 4, 36–50 (2015).
10. Chanoumidou, K., Mozafari, S., Baron-Van Evercooren, A. & Kuhlmann, T. Stem cell derived oligodendrocytes to study myelin diseases. *Glia* https://doi.org/10.1002/glia.23733 (2019).
11. van Bruggen, D., Agirre, E. & Castelo-Branco, G. Single-cell transcriptomic analysis of oligodendrocyte lineage progenitors during development. *Dev. Cell* 46, 504–517 e507 (2020).
12. Starost, L. et al. Extrinsic immune cell-derived, but not intrinsic oligodendroglial factors contribute to oligodendrogial differentiation block in multiple sclerosis. *Acta Neuropathol.* 140, 715–736 (2020).
13. Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* 89, 37–53 (2016).
14. van Bruggen, D., Agirre, E. & Castelo-Branco, G. Single-cell transcriptomic analysis of oligodendrocyte lineage cells. *Curr. Opin. Neurobiol.* 47, 168–175 (2017).
15. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386 (2014).
16. Wang, S. et al. Human iPSC-derived oligodendrocyte progenitor cells can myelinate and rescue a mouse model of congenital hypomyelination. *Cell Stem Cell* 12, 252–264 (2013).
17. Piao, J. et al. Human embryonic stem cell-derived oligodendrocyte progenitors remyelinate the brain and rescue behavioral deficits following radiation. *Cell Stem Cell* 16, 198–210 (2015).
18. Yeung, M. S. Y. et al. Dynamics of oligodendrocyte generation in multiple sclerosis. *Nature* https://doi.org/10.1038/s41586-018-0842-3 (2019).
19. Traiffort, E., Zakaria, M., Lauwrem, Y. & Ferrant, J. Hedgehog: a key signaling in the development of the oligodendroglial lineage. *J. Dev. Biol.* 4, https://doi.org/10.1186/s40009-018-0302-8 (2016).
20. Sluch, V. M. et al. Enhanced stem cell differentiation and immunopurification of genome engineered human retinal ganglion cells. *Stem Cells Transl. Med.* 6, 1972–1986 (2017).
21. Wang, Y., Wang, F., Wang, R., Zhao, P. & Xia, Q. 2A self-clipping peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci. Rep.* 5, 16727 (2015).
22. Daniels, R. W., Rossano, A. J., Macleod, G. T. & Ganetzky, B. Expression of multiple transgenes from a single construct using viral 2A peptides in Drosophila. *PLoS ONE* 9, e100637 (2014).
23. Sluch, V. M. et al. Highly efficient scarless knock-in of reporter genes into human and mouse pluripotent stem cells via transient antibiotic selection. *PLoS ONE* 13, e0201683 (2018).
56. Wang, J. et al. Robust myelination of regenerated axons induced by combined manipulations of GPR17 and microglia. Neuron https://doi.org/10.1016/j.neuron.2020.09.016 (2020).

57. Ellison, J. A. & de Vellis, J. Platelet-derived growth factor receptor is expressed by cells in the early oligodendrocyte lineage. J. Neurosci. Res. 37, 116–128 (1994).

58. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-labeling strategy to study cell-cell interactions in the mouse brain. Nat. Neurosci. 17, 794–803 (2014).

59. Ho, A. L. et al. PDGF receptor alpha is an alternative mediator of rapamycin-induced Akt activation: implications for combination targeted therapy of renal cell carcinoma. Cancer Res. 72, 4515–4525 (2012).

60. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201 (2015).

61. Lertsun, B. S. et al. MicroRNA expression profiling of oligodendrocyte differentiation from human embryonic stem cells. PloS ONE 5, e10480 (2010).

62. Wang, H. et al. miR-219 cooperates with miR-338 in myelination and promotes myelin repair in the CNS. Dev. Cell 40, 566–582.e565 (2017).

63. Dugas, J. C. et al. Dicer1 and miR-219 Are required for normal oligodendrocyte differentiation and myelination. Neuron 65, 597–611 (2010).

64. Zhao, X. et al. MicroRNA-mediated control of oligodendrocyte differentiation. Neuron 65, 612–626 (2010).

65. Suki, N. et al. Differentiation of oligodendrocyte precursors from Sox10–venus mice to oligodendrocytes and astrocytes. Sci. Rep. 7, 14133 (2017).

66. Zhao, C. et al. Dual regulatory switch through interactions of Tce1/tce4 with stage-specific partners propels oligodendroglial maturation. Nat. Commun. 7, 10883 (2016).

67. Ehrlich, M. et al. Rapid and efficient generation of oligodendrocytes from human induced pluripotent stem cells using transcription factors. Proc. Natl. Acad. Sci. USA 114, E2243–E2252 (2017).

68. Weng, Q. et al. Single-Cell transcriptomics uncovers glial progenitor diversity and cell fate determinants during development and gliomagenesis. Cell Stem Cell 24, 707–723.e708 (2019).

69. Craig, A. et al. Quantitative analysis of perinatal rodent oligodendrocyte lineage progression and its correlation with human. Exp. Neurol. 181, 231–240 (2003).

70. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. Prog. Neurobiol. 106–107, 1–16 (2013).

71. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. 27, 275–280 (2009).

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

73. Stuurt, T. et al. Comprehensive integration of single-cell. Data Cell 177, 1888–1902.e1821 (2019).

74. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for gene set enrichment analysis. Bioinformatics 23, 3251–3253 (2007).

75. Trapnell, C. Monocle: Differential Expression and Time-series Analysis for Single-cell RNA-Seq. http://cole-trapnell-lab.github.io/monocle-release/articles/v2.2.0/ (2016).

76. Fang, W., Xitz, C., Kallman, A., Ji, H. & Zack, D. Single-cell transcriptomic analysis reveals molecular diversity and developmental heterogeneity of human stem cell-derived oligodendrocyte lineage cells. Zenodo https://doi.org/10.5281/zenodo.4290971 (2020).

77. Kramer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 30, 523–530 (2014).

Acknowledgements
We thank Valentina Fossati (The New York Stem Cell Foundation Research Institute) for helping us establish the OPC differentiation protocol in the lab. This work was supported by grants from the Maryland Stem Cell Research Fund (MSCRF-1773 and MSCRFD-4317), Race to Erase MS, and NIH (P30 EY001765 and K99 EY029911), unrestricted funds from Research to Prevent Blindness, and generous gifts from the Guerrieri Family Foundation.

Author contributions
X.C. designed the study, conducted experiments, and wrote the manuscript. A.K. performed the Dropseq experiments, analyzed RNA-seq data, and edited the manuscript. W.F. analyzed the RNA-seq data, Z.I. and H.J. assisted with the analysis. C.B. assisted with experimental design and edited the manuscript. P.D. made the website. J.L.M. performed the pathway enrichment analysis. I.E.M.P., M.D.S., K.W., and P.A.C. performed the scRNA-seq work on the hiPSC-derived OPGs. P.A.C. also provided funding support. C.C. and H.-Q.M. provided nanofibers and assisted with the myelination assay. A.K., I.C., and T.-H.W. established the Dropseq platform in the lab. D.J.Z. designed the study, edited the manuscript, and provided funding support.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-021-20892-3.

Correspondence and requests for materials should be addressed to D.J.Z.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2021