Title: CelltypeR: A framework to identify and characterize cell types in human midbrain organoids using flow cytometry.

Authors: Rhalena A. Thomas[1], Julien M. Sirois[1], Shuming Li[1], Alex Gestin[1,2], Valerio E. Piscopo[1], Paula Lépine[1], Meghna Mathur[1], Carol X.Q. Chen[1], Vincent Soubannier[1], Taylor M. Goldsmith[1], Lama Fawaz[1], Thomas M. Durcan[1], Edward A. Fon[1]

Affiliations:
1.Early Drug Discovery Unit, Department of Neurology and Neurosurgery, Montreal Neurological Institute-Hospital, McGill University, Montreal, Quebec H3A 2B4, Canada
2. Université Paris-Saclay, Bures-sur-Yvette, France

Abstract

With the recent surge of single cell RNA sequencing datasets (scRNAseq) the extent of cellular heterogeneity has become apparent, yet it remains poorly characterized on a protein level in brain tissue and induced pluripotent stem cell (iPSC) derived brain models. With this in mind, we developed a high-throughput, standardized approach for the reproducible characterization of cell types in complex neuronal tissues. We designed a flow cytometry (FC) antibody panel coupled with a computational pipeline to quantify cellular subtypes in human iPSC derived midbrain organoids. Our pipeline, termed CelltypeR, contains scripts to transform and align multiple datasets, optimize unsupervised clustering, annotate cell types, quantify cell types, and compare cells across conditions. We identified the expected brain cell types, then sorted neurons, astrocytes, and radial glia, confirming these cell types with scRNAseq. We present an adaptable analysis framework providing a generalizable method to reproducibly identify cell types across FC datasets.

Introduction

Investigating the molecular, cellular, and tissue properties of the human brain requires the use of cellular models, as primary live human brains cannot be easily accessed for research. Patient-derived disease 3D tissues, such as human midbrain organoids (hMOs), derived from reprogrammed human induced pluripotent stem cells (iPSCs), provide a promising physiologically relevant model for human brain development and diseases, including neurodegenerative diseases such as Parkinson's disease[1–3]. As new models emerge, the complexity and reproducibility of these systems needs to be captured. To determine how faithfully organoids model the human brain and how organoids derived from individuals with disease differ from those derived from healthy controls, new approaches towards characterization are required. Effective and quantitative methods are needed to determine the cell types within these complex tissues and to apply these benchmarks reproducibly across experiments. At present, individual cells within brain or organoid tissue can be identified using single cell RNA sequencing (scRNAseq) or labeling of protein or RNA in tissue sections. These tools are useful but limited. scRNAseq is a powerful tool that has been used to identify known and novel cell types, cell states, and cell fate trajectories[4–6]. However, using scRNAseq to compare proportions or populations of cells between genotypes over multiple time points is not practical for hMOs and may result in sampling bias, as less than 1% of the whole tissue is sequenced. While scRNAseq provides detailed expression values to determine sub-types of cells, only relatively few samples can be run at a given time and all the cells must be alive and prepared in parallel, which can lead to technically challenging experiments. These experiments are also costly for the number of replicates needed to ensure enough power for comparing multiple time points, disease states, or pharmacological treatments[7–9]. Another option to quantify cell types is immunostaining or *in situ* hybridization of tissue sections. This has the advantage of capturing cell morphology and spatial resolution. However, sample preparation, image acquisition, and analysis are labour intensive and limited in quantitative accuracy. Moreover, for 3D tissues, either only a small section can be analyzed or the entire tissue must be reconstructed and only a few cell types can be detected at once[10,11].

Here, we use flow cytometry (FC) to measure the protein expression levels of a panel of cell surface markers enriched in specific brain cell types. FC is a fast, quantitative, and robust method, used widely in immunology and cancer research[12–14] but to date only sparsely in neuroscience. Typically in neurobiology, only two or three antibodies are used to distinguish between pairs of cell types[15,16] or to enrich one cell type[17,18]. Traditional FC analysis methods using FlowJo, the commercial FC analysis software package that is currently the standard in the field, are time consuming and subject to user error. Methods to standardize data preprocessing and analyze combinations of more than 3 antibodies in one experiment are starting to emerge[19,20]. However, no methods are available to automate cell type annotation in FC from complex tissues such as brain or 3D brain organoids using a large antibody panel. To create such an analysis framework, we produced an experimental dataset using cultured hMOs differentiated from human iPSCs[1,21,22]. Our workflow also provides the methods to select subtypes of cells and gate these cells for further analysis, such as RNAseq, proteomics, or enriching cultures. We select example cell populations, sort these cell types, and further characterize these with scRNAseq. Here we present a complete framework for annotating cell types within complex tissue and comparing proportions of cell types across conditions and experiments.

Results

An antibody panel to identify multiple cell types in human midbrain organoids

In **Figure 1A**, we provide a schematic of the CelltypeR analysis workflow (see methods) used to quantify and compare cell types from tissues with a complex mixture of cell types such as the brain. To test our CelltypeR pipeline, we used hMOs[22,23] differentiated from iPSC lines derived from three unrelated healthy individuals (**Table S1**). The hMOs were grown for 9 months in culture, a time point at which neurons are mature and myelination has been shown to occur.[1,24] Immunofluorescence staining of cryosections shows that these organoids contain neurons, astrocytes and oligodendrocytes (**Figure 1B**). In FC, combinations of the relative intensities of 2-3 antibodies are often used to distinguish between cell types. However, in hMOs we expect approximately nine cellular types with a continuum of stages of differentiation.[1,25,26] We first defined a panel of 13 antibodies, which includes well-characterized antibodies previously used to define neural stem cells, neurons, astrocytes, and oligodendrocytes or to define other cell types in cultured immortalized human cell lines, blood, or brain tissues (**Table S2**). We dissociated the mature hMOs and labeled the cell suspension with these antibodies then measured the fluorescence intensity values using FC. The single live cells were sequentially gated using FlowJo. The FC results show that each antibody has a range of intensities across different cells (**Figure 1C and S1**). We conclude that the antibody panel has the potential to be used to define cell types by identifying combinations of antibody expression profiles unique to different cell groups.

Validation of the antibody panel using 2D cultures and known cell type markers

To test the expression of the selected antibodies on known cell types, we separately differentiated iPSCs into dopaminergic neuronal precursor cells (DA NPCs), dopaminergic neurons (DA neurons), astrocytes, and oligodendrocytes (oligos) (**Figure 2A and Table S3**). The cultures were dissociated, and the 13-antibodies in the FC panel were applied. We examined the staining for each antibody across the cultured 2D cells (**Figure 2B**). Within each cell type there was a variation in protein levels that could be used to define subgroups of cells. To identify subtypes of cells and visualize the markers, we applied unsupervised clustering developed as part of the CelltypeR workflow. Some tools exist for automated processing and formatting of FC and numerous tools exist for cluster analysis of single cell transcriptomic data that can be applied to other FC data. Thus, we took advantage of these existing tools and created new functions in an R package to process FC data (see methods). We combined the

FC acquired antibody intensities from the five separate iPSC derived cultures, normalized the data, performed dimensional reductions (PCA) and used Louvain network detection to identify groups of cells. The UMAP visualization shows separate groups for each of the five cell types with some overlap (**Figure 2C**). We observe that the iPSCs are mostly separated from all the other cell types. The DA NPC culture splits into separate groups and overlaps with different cell types. An isolated population of cells forms DA NPCs, some of the culture has started differentiating into early neurons, a small proportion are differentiating into astrocytes, and a small group is consistent with neural stem cells. The oligodendrocyte culture splits into two groups: the true oligodendrocytes expressing the marker O4, and radial glia indicated by high expression of both neuronal and glial markers. We conclude from these findings using iPSC-derived 2D cultures that our antibody panel can distinguish different cell types and subgroups of cell types that we expect to find in 3D hMOs and other complex neuronal tissues.

Identification of different brain cell types in human midbrain organoids

To identify cell types within hMOs using the antibody panel, we ran our R preprocessing pipeline to align and normalize the data. To compare samples from different iPSC lines, different batches of hMOs, and measurements run on different experiment days, we developed methods to combine and harmonize samples, which is the first step in the computational pipeline. We combined nine hMO samples and selected a subset of the total cells (9000 cells or the max number of cells available). The samples were first merged, then transformed and aligned to reduce batch effects and finally retro-transformed for better cluster visualization (**Figure S2**). If removing batch effects is not desired (as in the separate cell cultures above), the preprocessing is stopped after merging. The hMOs are expected to contain a combination of neurons, neural precursor cells (NPCs), astrocytes, oligodendrocyte precursors (OPCs), oligodendrocytes, radial glia (RG), stem cells, pericytes, endothelial cells, and epithelial cells, all differentiated from the starting iPSCs. The standard method of manually defining cell groups using FlowJo or multiple scatter plots in R is time consuming and not reproducible across experiments. To overcome this barrier, we developed tools to identify cell types described below: A) A correlation cell type assignment model (CAM) using a custom reference matrix and B) clustering parameter exploration functions with tools to visualize and summarize of protein expression levels.

We created a reference matrix with the predicted relative expression of each cell surface marker in different cell types expected to be present in hMOs based on known brain cell types and previous hMO scRNAseq. Using scRNAseq data from human brain and organoids, total mRNA on brain cell types, and FC (Figure 2), we calculated the relative expression levels for each marker in our antibody panel (**Figure 3A**). Our CAM function calculates the correlation of protein expression levels of the 13 markers in each hMO-derived cell to the expression levels of the same markers in the reference matrix we created, calculating the Pearson correlation coefficient, R. The R value is calculated for each cell type in the reference matrix and one cell type out of the nine possible cell types is assigned for a given hMO derived cell (**Figure 3B and S3**). To avoid false cell type assignments, we added a correlation coefficient threshold of 0.45, where hMO derived cells with R values below the cut-off are assigned as ‘unknown’. Many hMO-derived cells have the highest correlation with oligodendrocytes but are labelled as ‘unknown’ because of the applied threshold. Some hMO-derived cells correlated highly with two cell types. When this was the case, these cells were assigned a merged cell type, and may represent an intermediated cell type (**Figure 3C and S3-5)**. The most common cell pairs of cell types with a first and second top correlation within 0.05 are similar cell types, the most frequent pair is neurons and NPCs, which are the same cell type on a continuum of differentiation (**Figure 3C and S5**). The most frequent assignment is ‘unknown’ cell type, indicating that these cells didn’t correlate highly with any of the predicted cell types expression patterns (**Figure S6**). Clustering accounts for the problem of ‘unknown’ predictions because similar cells are grouped together. We created functions to identify the topmost predicted cell types per cluster, so by ignoring ‘unknown’ we can conclude the second most abundantly predicted cell type is the main cell type of a given cluster (**Figure S6**). We also applied the correlation assignment to the 2D culture data and found about half of the cells are correctly predicted in each different cell culture (**Figure S7**). Although correlation assignment is a useful tool to provide biologists with a predicted cell type, it does not deliver the accuracy needed to quantify cell types across experiments and therefore must be used in combination with other methods.

Using the functions in our CelltypeR library we performed unsupervised clustering using Louvain network detection and visualized the protein expression levels (**Figure S8-9**). Clusters were annotated with cell types using a combination of marker expression by cluster and the output from the correlation predicted cell types (**Figure 3D**). We identified astrocytes, radial glia (RG), epithelial cells, endothelial cells, NPCs, neurons, a small proportion of oligodendrocytes, and stem cell-like cells in the hMOs. Some cells have low relative antibody expression and form a cluster together and these cells were annotated as 'unknown' (**Figure 3E**). Another cluster has low expression overall, but some expression of markers indicating a mix of glial cells and neurons. This cluster was annotated as 'Mixed'. Clustering the hMO cells identified distinct subpopulations of RG, astrocytes, and neurons. These populations can be broken into further subgroups (**Figure S10**). We conclude that our workflow can be used to annotate cell types in hMOs.

Comparison of cell types between iPSC lines and hMO batches

After annotating a subset of 9000 cells from each of the nine hMO samples, we next analyzed the total available cells (**Table S1**). We again followed the CelltypeR workflow, but now we used the labelled subset of cells to annotate the full dataset. Using the subset of annotated cells, we trained a random forest classifier model (RFM) (**Figure S11)** and then applied those labels to the complete dataset to predict the type of each cell. We clustered the full dataset and visualized the predicted cell type annotations using RFM, CAM and Seurat[27] label transfers (**Figure S12**). To annotate the cells in the full dataset from the nine MOs, we applied the CelltypeR tools using three methods of cell type prediction and inspection of expression levels in each cluster in UMAP and heatmap visualization (**Figure S13**). We observe the same cell types in the full dataset as in the subset of data; however, we now identify a tiny cluster of OPCs, and more stem cell-like cells (**Figure 4A**). Using the cell type predictions from the RFM and the Seurat label transfers we now have an indication of the radial glia cells within the 'unknown' and 'mixed' clusters (**Table S4**). Based on the visualization of markers, the 'mixed' cluster has cells of glial lineage, and the unknown cluster has cells with a neuronal lineage (**Figure 4A**). Subgroups of neurons and glia are clearly defined by different expression patterns of the antibody panel (**Figure 4B**). In the full dataset we observe more subgroups of the main cell types (**Figure S14**).

Visualizing the distribution of cell types in hMOs derived from each cell line, we can see there are some differences in the proportion of cell types (**Figure 4C, D**). Differences are also observed for the other variables, namely days in culture and experiment date, but no differences were observed between the two batches. This indicates that there is low variation between batches of hMOs (**Figures S15**). We next did proportionality tests to determine if the differences in cell types between the cell lines are significant. The proportion of neurons 1 and some of the glial populations are increased while the proportions of neurons 2, oligodendrocytes, and stem cell-like cells are decreased in the AIW002 line compared to the AJG001 and 3450 cell lines (**Figure 4D and S16**). We created functions to compare the mean surface marker expression between different variables within different cell type populations (**Figures 4E**). We also built functions in our R package to run ANOVAs, post-hoc tests, and identify significant differences. We tested if expression markers between groups are significantly different between cell lines, the number of days hMOs are in culture, hMO batch, and experiment day across cell types using one-way ANOVAs and found some significant differences (**Table S5**). We next performed two-way ANOVAs with marker expression and cell line followed Tukey's post-hoc tests. There are significant differences in overall marker expression levels between the AIW002 and 3450 lines in epithelial cells, AJG001C and both 3450 and AIW002 in neurons 1, 3450 and both AJG001C and AIW002 in neurons 2, and AIW002 and AJG001C in oligodendrocytes (**Table S6**). We ran the same statistics to compare marker expression levels between different amounts of time spent in culture and observed significant differences in some cell types (**Table S5)**. However, pair-wise comparisons show that these differences likely reflect the differences between experiment dates because hMOs at 263 and 284 days in culture measured on the same day do not show any significant differences (**Table S7**). Individual marker expression between cell lines, days in culture, batch, or experiment date are not significant in very many cases (**Figures S17-20**). Using our framework, we can reliably quantify cell types and compare proportions of cells and levels of antibody expression across different conditions. We find significant differences in the proportion of cell types and in marker expression levels within cell types between different healthy control iPSC lines.

Isolating populations of interest identified by CelltypeR clustering analysis

After annotating the dataset, we could plot the proportion and mean expression of each antibody marker in each group to try and define the relative marker expression of a given cell group and then isolate that population by FACS. However, manually reverse engineering a gating strategy is difficult with more than a few cell type markers. Thus, we defined cell types using CelltypeR, applied the package hypergate[28] to identify which combinations of antibody markers clearly define a given cell population, and then manually gated these cells in FlowJo (**Figure 5A**). The gating accuracy for all cell types is above 95% **(Table S8 and S9)**. We next followed the CelltypeR workflow using the newly generated gated files to annotate the cells in the FlowJo gated populations (**Figure 5B**). The most frequent CelltypeR annotated cell type within each gated population is the intended cell type, except for NPCs, where Neurons 1 is the most common cell type (**Figure 5C and Table S10**). We find that CelltypeR can define cell types and gates and these can be used to effectively gate the desired cell types.

Analysis of FACS sorted neuronal and glia populations followed by single cell sequencing analysis.

Our workflow can be used to enrich populations of interest by FACS sorting selected populations for further analysis. We selected four cell types: neurons 1, neurons 2, astrocytes, and radial glia. We then designed a gating strategy to simultaneously sort the four populations (**Figure 6A**). We sorted the hMO-derived cells using the defined gates, split the samples, and then acquired FC measurements and scRNAseq on the sorted populations. The protein expression levels in the sorted populations match the expected levels from the gates (**Figures 6B**). We also obtained a single cell transcriptomic library for each of the FACS sorted populations (see methods). We first compared the RNA expression levels of the genes corresponding with the protein expression levels measured by FC and found they highly correlate (**Figure 6C and Table S11**). The four populations were merged, clustered, and plotted on a UMAP to visualize the overlap between the different sorted cell types (**Figure 6D**). The *Neurons1* population is mostly separate from the other populations with some overlap with *Neurons2*. Clusters were first annotated for main groups of cell types: DA neurons, neurons, NPCs, radial glia, and astrocytes. These main cell types were subset and annotated for subtypes of cells using differential gene expression between clusters (**Figure 6E, S21-24 and Table S12)**. Next, we calculated the proportion of cellular subtypes in the FACS sorted populations (**Figure 6F and Table S13**). We found that non-DA neurons in *Neurons1* are excitatory and mature neurons as well as NPCs and ventral zone (VZ) radial glia undergoing neurogenesis. The non-DA neurons in the *Neurons2* population are GABAergic, serotonergic (5HT), and neurons with potential to be reactivated as neural stem cells. As quantification of DA neurons is of particular interest in hMO for Parkinson's Disease, we find that Neurons1, Neurons2, and RadialGlia all contain DA neurons. The Neurons1 FACS population has slightly more DA neurons overall, specifically the Substantia Nigra (SN) subtype, whereas the Neurons2 FACS population has more of the ventral mesencephalon (vm) subtype (**Figure S25 and Table S14)**. The two FACS sorted neuron populations contain distinctive subtypes of DA and non-DA neurons. The astrocyte population split into three subgroups, immature, resting and reactive. The radial glia population contains five different subtypes (**Figure 6E, S24 and Table** S12). We show that each FACS sorted population is enriched in the expected cell type and there are identifiable subtypes within these groups, confirming the effectiveness of the celltypeR framework.

Discussion

Taken together, we present the first protein based complete workflow to identify, quantify, and compare cell types in complex 3D tissue, specifically hMOs. We define a 13-antibody body panel that can be used to distinguish between eight different brain cell types and identify subtypes of astrocytes, radial glia, and neurons. The panel is modular and can be altered or expanded and will still function with the computational workflow. In our CelltypeR library, we provide a method to preprocess and merge FC samples, acquired from multiple samples at different dates. We created tools to optimize and visualize clustering and to assist in consistent cell type annotation. We also created functions to quantify cell types and compare different conditions. The same workflow with sorting can be used to isolate a more homogenous subpopulation of a given cell type to perform other assays such as proteomics or lipidomics, or to replate the cells in culture to grow as a purified population. Here we selected four populations, FACS sorted the cells, and then performed scRNAseq analysis. We confirmed that each of the populations, *Neurons1*, *Neurons2*, *RadialGlia* and *Astrocytes*, are all highly enriched in the expected cell types. Further analysis of the scRNAseq data identified subtypes within each cell type group. We identified DA neurons within both neuronal populations but find different DA neuron subtypes are more enriched in the two FACS sorted neuron populations. We also identified TPGB as a DA subtype marker (ventral), in agreement with a recent publication proposing TPGB as a marker of ventral DA neurons in mice.[29]

In our analysis of the differences between three healthy control iPSC lines, we find a clear difference in the proportion of cells for the two subtypes of neurons between AIW002 compared to the other two lines, AJG001 and 3450. AIW002 has more *Neurons1* with high CD24 expression and fewer of the *Neurons2* population, with lower CD24 expression than AJG001 and 3450. scRNAseq reveals the *Neurons1* population has more NPCs and DA neurons. We also find that AIW002 has more radial glia, fewer astrocytes, and fewer oligodendrocytes than the other two lines, indicating this cell line may be less mature. AIW002 might mature at a slower rate or given the very late age of the organoids, maintain a less mature state perpetually. These findings also indicate that to study the role of myelination, the AJG001 or 3450 lines could be a better choice than AIW002.

The CelltypeR workflow we present can be applied for developmental experiments to track the emergence of neurons and mature glia populations and the loss of stem cells over time. Furthermore, cell types in hMO disease models derived from patient iPSCs can be compared to control hMOs by quantifying cell types over time. In the current cell surface panel, most cells in our hMO data are easily annotated. However, some cells are not easily identified by the FC panel. It is possible these cells are not expressing many proteins at the cell surface or that these cells represent a cell type not well covered by the antibody panel. Within our workflow the antibody panel can be easily changed. Our starting antibody panel could be fine-tuned and tested in our workflow. Furthermore, new panels appropriate for different complex tissues, for example kidney or gut, can be designed to distinguish cell types using the CelltypeR workflow. For changes in the antibody panel, a reference matrix from experimental or public data needs to be created to use the correlation prediction method. We have also outlined all the steps needed for creating a reference matrix. Altogether, we have created an adaptable method to reproducibly identify and quantify cell types in complex 3D tissues using an FC panel. We developed a novel scalable single cell biology workflow to quantify cell types quickly and efficiently in complex neural tissues, specifically hMOs, across multiple replicates and experimental conditions.

Acknowledgments

T.M.D. received funding through the McGill Healthy Brains for Healthy Lives (HBHL) initiative, the CQDM Quantum Leaps program with support from Brain Canada, the Alain and Sandra Bouchard Foundation, the Sebastien and Ghislaine Van Berkom Foundation, the Chamandy Foundation and the Mowafaghian Foundation. T.M.D is supported by a project grant from CIHR (PJT-169095).

E.A.F. is supported by a CIHR Foundation grant (FDN – 154301), a Fonds d'Accéleration des Collaborations en Santé (FACS) grant from CQDM/MEI and by a Canada Research Chair (Tier 1) in Parkinson's disease.

R.A.T. received funding through the McGill Healthy Brains for Healthy Lives (HBHL) Postdoctoral Fellowship and Molson NeuroEngineering Fellowship. The authors thank David Kalaydjian and Nguyen-Vi Mohamed for helping to test early midbrain organoid differentiation protocols. We thank Dr. Jo Anne Stratton for use of the 10X chromium controller and for constructive feedback on the manuscript.

Contributions

The project was conceived by RAT and JMS. The production and maintenance of hMOs was done by MM and PL. FACS sorting, panel optimization, antibody titration and flow acquisition were done by JMS. Sample preparation for flow acquisition, FACS sorting and scRNAseq were performed by JS, RAT and PL. The iPSC, NPC and DA neuron cultures and corresponding IF by done by CXQC. Astrocyte cultures and IF by VS. Oligodendrocyte and OPC cultures and IF. The hMO cryosections were prepared by PL and the IF was done by VEP. The scRNAseq library preparation was done by TMG and LF. FlowJo analysis was performed by JS and RAT.
Computational workflow and the R library were designed, tested, and managed by RAT. R functions were written by AG, SL and RAT. Computational analysis was performed by RAT. Data was interpretation by RAT, TMD and EAF. The project supervised by RAT, TMD and EAF. The manuscript was written by RAT with contribution for selected sections by JMS, TMG, EAF, TMD.

Methods

1. iPSC lines used for hMOs

The three iPSC cell lines were used: AJG001C, AIW002 and 3450. All were previously reprogrammed from peripheral blood mononuclear cells as previously described.[21] All work with human iPSCs was approved by McGill University Faculty of Medicine and Health Sciences Institutional Review Board (IRB Internal Study Number: A03-M19-22A).[21]

2. Cell culturing conditions

2.1 2D cultures and differentiation.

The control cell line AIW002 was used for all 2D cell cultures. AJG001C, AIW002 and 3450 were used for hMOs. Prior to differentiation, the iPSC cultures were maintained and expanded on Matrigel coated plates and grown in either mTeSR1 or E8 media as previously described.[21,30]

Dopaminergic neural precursor cell (DA-NPC) cultures were generated by dissociating iPSCs into single cell suspensions and then culturing these cells in low attachment plates to generate embryoid bodies (EBs).[31] EBs were re-plated onto polyornithine and laminin-coated plates and differentiated into neural rosettes, which were then differentiated into DA-NPCs. DA neurons were differentiated from DA-NPC cultures on laminin coated culture flasks in neural basal media with supplements and inhibitors as described.[32]

To derive oligodendrocyte precursor cells (OPCs) and oligodendrocytes we used a three phase protocol as previously described.[33,34] In phase one, iPSCs were induced towards neural progenitors while being patterned with Retinoic Acid in order to resemble spinal cord progenitors. The Sonic Hedgehog pathway was activated for ventral patterning to recapitulate the conditions of the oligodendrocyte fate. The progenitors were subsequently expanded as EBs with the addition of the bFGF. In phase two, OPCs were expanded in suspension and subsequently plated onto polyornithine/laminin-coated vessels for adhesion. Growth medium mitogens were added for differentiation and maintenance of the OPCs. PDGRF positive images were acquired at this phase. In phase three, mitogens are withdrawn to allow the progenitors to exit the cell cycle and to complete differentiation into myelinating oligodendrocytes. Imaging and FC were performed in this phase when oligodendrocytes would generate O4 positive cells.

Astrocytes were derived from NPCs cultures as previously described.[35] NPCs were seeded at low cell density and grown in NPC expansion medium. The next day, medium was replaced with ‘Astrocyte Differentiation Medium 1’. Cells were split 1:4 every week and were maintained under these culture conditions for 30 days. At DIV50, cultures were switched to ‘Astrocyte Differentiation Medium 2’ and maintained with half medium changes every 3-4 days.

2.2 Human midbrain organoids

hMOs (AJG001C, AIW002 and 3450) were derived from iPSCs cultures according to the established protocols.[23,36] For each healthy control iPSC line, iPSCs were seeded in separate ultra-low attachment plates in neural induction medium for EBs to form. On day four, medium was changed to midbrain pattering medium to promote a dopaminergic neural cell fate. On day seven, hMOs were embed in Matrigel. On day eight, hMOs were transferred to 6-well plates with 4-6 hMOs per cell line in organoid growth media and placed in shaking cultures. hMOs were maintained in shaking cultures with media change every 2-3 days.[36] In the hMO samples used for gating and sorting neuronal and glia populations a newer protocol was used.[23] Dissociated iPSCs were seeded in eNuvio disks for EB formation and Matrigel embedding, then transferred to bioreactors for culture maintenance. Media changes were performed weekly and all the same growth mediums were used in both protocols.

3. Immunofluorescence

3.1 iPSC, NPCs and Dopaminergic Neurons

Cells were fixed in 4% PFA/PBS at RT for 20 minutes, permeabilized with 0.2% Triton X-100/PBS for 10 min at room temperature (RT), and then blocked in 5% donkey serum, 1% BSA and 0.05% Triton X-100/ PBS for 2h. Cells were incubated with primary antibodies: MAP2 (1:1000, EnCor Biotech CPCA-MAP2); Nestin (1:500, Abcam ab92391); SSEA-4 (1:200, Santa Cruz Biotechnology sc-21704); in blocking buffer overnight at 4 °C. Secondary antibodies were applied for 2h at RT, followed by Hoechst 33342 (1/5,000, Sigma) nucleic acid counterstain for 5 minutes. Immunocytochemistry images were acquired using Evos FL-Auto2 imaging system (ThermoFisher Scientific).

3.3 Astrocytes.

Cells were fixed 15 minutes at room temperature with 4% formaldehyde in PBS, followed by 3 washes of 5 minutes in PBS. Cells were permeabilized for 10 min at RT in blocking solution: 5% normal donkey serum (JacksonImmunoResearch Laboratories, West Grove, PA), 0.1% Triton-X-100, and 0.5 mg/ml bovine serum albumin (Sigma-Aldrich) in PBS. Cells were incubated for 1h at RT before overnight incubation at 4°C with primary antibodies: Glial Fibrillary Acidic Protein (GFAP) (1/500 Dako Cat. Number Z0334); AQP4 (1/500, SIGMA, cat# HPA014784). Secondary antibodies were incubated 2h at, followed by Hoechst 33258 (1/5,000, Sigma) for 5 min, mounted with Fluoromount-G, and examined by fluorescence microscopy.

3.4 Oligodendrocytes and OPCs.

Cells were fixed in 2% PFA for 10 min and blocked in 5% BSA, 0.05% Triton for 1h. Mouse anti-O4 (R&D, MAB1326) was added in live cells before fixation for 1h at a final concentration of 1µg/mL. Rabbit anti-PDGFRa (Cell Signaling, 3174) was added post-fixation at a dilution of 1:200 and incubated overnight at 4°C. Secondary antibodies were added at a dilution of 1:500 and incubated for 2h at RT. Nuclei were identified with incubation with Hoechst 33342 (1/5,000, Sigma) for 5 min.

3.5 Midbrain organoids

hMOs were washed in PBS and then fixed for 2hours in 4% PFA diluted in PBS at RT, then placed in a sucrose gradient overnight at 4°C. hMOs were then embedded in Optimal Cutting Temperature Compound (OTC) (Fisher Healthcare 23-730-571) and frozen. Cryosections of 20mM were cut using Cryostat Cryostar NX70 (Thermo Scientific). The slides with the sections were washed 2 times in ddH2O to remove the OCT, permeabilized 20min in 0.1% Triton-PBS and blocked for 1h in 5% Normal Donkey Serum (Jackson Immuno Research Laboratories, West Grove, PA), 0.2% Triton, 0.5mg/mL BSA (Sigma-Aldrich) in PBS. Primary antibodies: anti-O4 (1:200, R&D, MAB1326); Glial Fibrillary Acidic Protein (GFAP) (1/500 Dako Cat. Number Z0334); and MAP2 (1:1000, EnCor Biotech CPCA-MAP2) were diluted in blocking solution and incubated at RT for 1h. Fluorescent-labeled secondary antibodies (Invitrogen) were added at a dilution of 1:500 and incubated for 45min. Nuclei were identified with Hoechst 33258 (1:5000, Sigma). Cover slides were mounted using Fluoromount mounting medium (Sigma-Aldrich) and imaged using confocal microscopy (Leica TCS SP8 confocal).

4. Sample preparation for flow cytometry

4.1 Tissue dissociation and processing – Main data set hMOs

hMOs were dissociated with a combination of enzymatic digestion and mechanical dissociation. First, three individual hMOs from each of the data set of nine samples were removed from shaking cultures and combined into one 15mL tube. Pooled hMOs were washed three times with Dulbecco's PBS (D-PBS) (Wisent) to completely remove remaining culture media. Then, after completely removing D-PBS, 2mL of TrypLE express (without phenol red) (ThermoFisher) was added to each sample. The hMOs were incubated at 37°C for ten minutes then removed to be subjected to mechanical dissociation by pipette trituration (slowly pipetting up and down ten times). The incubation and the pipette trituration are repeated twice more. Afterwards, 8mL of D-PBS was added to the samples to stop the enzymatic reaction. The samples were filtered through a 30µm filter (Miltenyi Biotec) to remove any clumps remaining after digestion and dissociation. Samples were washed twice more with D-PBS.

4.2 Tissue dissociation and processing – Sorting data set hMOs

hMOs were dissociated with a combination of enzymatic digestion and mechanical dissociation. First, twenty individual hMOs were removed from a bioreactor and combined into one 50mL tube. Pooled hMOs were washed three times with Dulbecco's PBS (D-PBS) (Wisent) to completely remove remaining culture media. Pooled hMOs were transferred to a gentleMACS M-Tube (Miltenyi Biotec). Then, after completely removing D-PBS, 2mL of TrypLE express (without phenol red) (ThermoFisher) was added to each sample. The hMOs inside the M-Tube are then next placed on an automated GentleMACS Octo Heated dissociator. The settings for the dissociation were as follows: 37°C is ON. Spin -20rpm for 24 minutes. Spin 197rpm for 1 minute. After incubation, 8mL D-PBS was added to the samples to stop the enzymatic reaction. The samples were filtered through a 30µm filter (Miltenyi Biotec) to remove any clumps remaining after digestion and dissociation. The samples were then washed twice more with D-PBS.

4.3 Tissue dissociation and processing – 2D cell cultures

T-flasks containing cells were washed in PBS then incubated at 37°C in 2mL of TrypLE express (without phenol red) (ThermoFisher) for 5-20 minutes depending on cell type. Cells were washed off the growth surface with a pipette, then manual dissociated by trituration until no clumps were seen and transferred to a 15ml tube. Cells were washed twice in D-PBS.

4.4 Antibody staining – All samples

After counting and isolating one million cells, single cell suspensions were incubated for 30 minutes at room temperature in the dark with Live/Dead Fixable dye to assess viability. Single cell suspensions were washed twice with D-PBS to remove any excess dye. After, single cell suspensions were incubated for 15 minutes at room temperature in the dark with Human TruStain FcX (Biolegend) at a concentration of 5µL per million cells to block unspecific Fc Receptor binding. Single cell suspensions were washed once with FACS Buffer (5% FBS, 0.1% NaN_3 in D-PBS) and then incubated for 30 minutes at room temperature in the dark with a fluorescence-conjugated antibody cocktail in FACS Buffer (Methods Table 1). The information regarding working dilutions used in this antibody cocktail is in Methods Table 1. The optimal working dilutions were determined by titrations with similar hMOs and experimental conditions. After incubation, single cell suspensions were washed twice with FACS Buffer and resuspended in FACS Buffer. Samples were placed at 4°C until ready to be analyzed by flow cytometry.

In parallel, compensation control staining was performed with the same conditions as the single cell suspensions. The compensation controls used are UltraComp eBeads™ Plus Compensation Beads

(ThermoFisher) and ArC™ Amine Reactive Compensation Bead Kit (ThermoFisher) Samples were placed at 4°C until ready to be acquired by flow cytometry.

Table 1: Panel of antibodies and protein targets to measure using Flow Cytometry to identify cell types in hMOs with Fluorochrome and antibody information indicated.

Antibody	Gene/Protein Name	Fluorochrome	Ab clone	Manufacturer (CAT#)	Dilution used
CD44	CD44	BV421	BJ18	Biolegend (338810)	1:192
Aquaporin-4	AQP4	AF488	Polyclonal	Bioss (bs-0634R-A488)	1:28
GLAST	GLAST	APC	ASCA-1	Miltenyi Biotec (130-095-814)	1:20
HepaCAM	HepaCAM	AF594	Polyclonal	Bioss (bs-5840R-A594)	1:333
CD71	TNR	BV650	CY1G4	Biolegend (334116)	1:333
CD184	CXCR4	BV605	12G5	BD Optibuild (334116)	1:48
CD133	PROM1	PE-Cy7	clone 7	Biolegend (372810)	1:333
CD15	FUT1/SSEA-1	BV785	W6D3	Biolegend (323044)	1:48
CD29	ITGB1	APC-Cy7	TS2/16	Biolegend (303014)	1:48
CD56	NCAM	PercP-Cy5.5	5.1H11	Biolegend (392420)	1:96
CD24	CD24	BV711	ML5	Biolegend (311136)	1:192
O4	O4	PE	REA576	Miltenyi Biotec (130-117-357)	1:31
CD140α	PDGFRalpha	AF700	PRa292	R&D Systems (FAB1264N)	1:40

4.5 Flow Cytometry acquisition – All data sets

Single cell suspensions were acquired on an Attune NxT (ThermoFisher). The information for the configuration of this Flow Cytometer is in Methods Table 2. Daily CS&T performance tracking was done prior to cell acquisition by recommendation of manufacturer. PMT voltages were determined by Daily CS&T performance tracking. Compensation controls were also acquired, creating an acquired compensation matrix. Between 48 000 to 338 000 cells were acquired per sample.

4.6 Flow Cytometry cell sorting defined by CelltypeR workflow – Sorting data set

Single cell suspensions were sorted on a FACSAria Fusion (Becton-Dickinson Biosciences). The information for the configuration of this Flow Cytometer is in Methods Table 2. Daily CS&T performance tracking was done prior to cell acquisition by recommendation of manufacturer. PMT voltages were determined by Daily CS&T performance tracking. Compensation controls were also acquired, creating an acquired compensation matrix.

Table 2: ThermoFisher’s Attune NxT optical path configuration.

Laser	Wavelenght (nm)	Detector	Dichroic mirror	Filter (Band pass)
Violet	405	VL6	740 LP	780/60 BP
		VL5	680-740	710/50 BP
		VL4	635-680	660/20 BP
		VL3	555-635	610/20 BP
		VL2	495-555	525/50 BP
		VL1	417-495	450/40 BP
Blue	488	BL2	555 LP	695/40 BP
		BL1	503-555	530/30 BP
		SSC	-	488/10 BP
Yellow-Green	561	YL3	650 LP	780/60 BP
		YL2	600-650	620/15 BP
		YL1	577-600	585/16 BP
Red	640	RL3	740 LP	780/60 BP
		RL2	690-740	720/30 BP
		RL1	654-690	670/14 BP

Methods Table 3: BD's FACSAria Fusion optical path configuration.

Laser	Wavelenght (nm)	Detector	Dichroic mirror (Long pass)	Filter (Band pass)
Violet	405	A	750 LP	780/60 BP
		B	690 LP	710/50 BP
		C	630 LP	660/20 BP
		D	595 LP	610/20 BP
		E	505 LP	525/50 BP
		F		450/50 BP
Blue	488	A	655 LP	695/40 BP
		B	502 LP	530/30 BP
		C		488/10 BP
Yellow-Green	561	A	735 LP	780/60 BP
		B	685 LP	710/50 BP
		C	630 LP	670/14 BP
		D	600 LP	610/20 BP
		E		582/15 BP
Red	640	A	755 LP	780/60 BP
		B	690 LP	730/45 BP
		C		670/30 BP

5. Single cell sequencing of FACS sorted populations

Three separate tubes of AIW002 hMO were dissociated as described above. At the antibody labelling stage oligonucleotide tagged antibodies (Hashtags, Biolegend) were added with the other cell type specific antibodies. The cells were sorted into FACS buffer. The same sorted populations from each of the three samples (replicates) were combined after sorting. These four populations were sorted into four gates and were sorted until the sample with fewest cells (Neurons1) contained 100,000 events. The sorted samples were centrifuged for 5minutes at 400g and resuspended in 250 ml of D-PBS + 0.1% BSA. The cell concentrations were calculated with FACSAria Fusion (Becton-Dickinson Biosciences). The single cell suspensions were diluted to 1000 cells/ml targeting ~15,000 cells captured for sequencing. One sample was prepared for each FACS sorted population.

Following the creation of the cell suspension, the Chromium NextGEM Chip G (PN-1000120) was then loaded as per manufacturer recommendation and run on the Chromium Controller (PN-1000204) for GEM creation. All proceeding thermocycler steps in the 10X protocol were carried out on a Bio-Rad C1000 Touch thermal cycler (1851196). Following GEM-RT incubations, samples were stored at 4°C overnight. Post GEM-RT cleanup and cDNA amplification were carried out per manufacturer protocol. Samples were stored at -20°C until they were processed for library generation. 3' gene expression and cell surface protein libraries were constructed per manufacturer protocol and stored at -20°C until sequencing submission. 25 mL of each sample library was sent for sequencing at the McGill Genome Centre.

Methods Table 4: Single cell RNA sequencing reagents

Product	Vendor	Product number
Chromium Next GEM Single Cell 3’ kit v3.1	10X Genomics	1000268
3’ Feature Barcode kit	10X Genomics	1000262
Chromium Next GEM Chip G Single Cell Kit	10X Genomics	1000120
Dual Index Kit TT Set A	10X Genomics	1000215
Dual Index Kit NT Set A	10X Genomics	1000242
Dynabeads MyOne SILANE	10X Genomics	2000048
Magnetic Separator	10X Genomics	230003
SPRIselect Reagent	Beckman Coulter	B23318
Qiagen Buffer EB	Qiagen	19086

Data processing

6.01 Flow Cytometry data cleanup for analysis – All data sets

The data generated was cleaned up using FlowJo (version 10.6) (Becton-Dickinson Biosciences). Briefly, a starting gate was used to select appropriate cell size (X: FSC-A, Y: SSC-A). A second gate was used to discriminate doublets from the analysis (X: FSC-W, Y: FSC-H). Finally, the last gate was used to remove dead cells from the analysis (X: LiveDead Fixable Aqua, Y: FCS-A). See Methods Figure 1 for a gating example. After data cleanup, a new .fcs file was generated with FlowJo and exported for further analysis done with R.

6.02 Data analysis and CelltypeR R library

All computations were performed in R. We created a R library of functions to perform the analysis, CelltypeR. Our functions required functions from multiple other R libraries referenced in descriptions to follow.
The R library can be found, along with workbooks for the complete workflow and generation of each figure, at https://github.com/RhalenaThomas/CelltypeR_single_cell_flow_cytometry_analysis

Computational Workflow:

1. Data preprocessing:
    a. Read FlowJo files into R.
    b. Create a data frame with intensity measurements for each marker for all samples within the experiment to be analyzed.
    c. Harmonize data if desired.
    d. Create a Seurat single cell object for further analysis.
2. Creation of cell type clusters
    a. Clustering optimization to compare clustering methods and parameters and visualize results.
    b. Summarize statistics to compare clustering methods and parameters.
    c. Select one method and smaller parameter space to compare cluster stability.
    d. Evaluate statistics and visualization to determine the best clustering method for a given visualization.
3. Cluster annotation

    a. For first data set: marker visualization and correlation assignment model.
    b. For subsequent data sets: marker visualization and correlation assignment model, Random Forest Model, Seurat Label Transfer.
4. Quantify cell types and measure expression levels of markers within cell types.
5. Define marker levels for cell types.
6. Statistical analysis between different groups of interest.

6.03 Data preprocessing

The .fcs files without dead cells, debris, and doublets created in FlowJo are read into R and processed. The .fsc files contain area, width, and height of the fluorescence signal for each marker as well as the forward and side scatter of the light. Then R using the flowCore package is used (Hahne et al., 2009). The area values for each channel are selected to represent the expression intensity for each antibody. All the .fsc files within one folder are read into into one R data object. A dataframe is created with the channels and saved for further use. Individual cell cultures and hMO organoid samples for testing the pipeline and gating were used in this raw format to create a Seurat single cell data object.

For the hMO samples, the data was aligned to remove batch effects and technical variability. Each file represents an experimental sample, and the samples were aligned as follows: First, to enhance the distinction between positive and negative antibody staining the raw data is transformed using the biexponential transform function from flowCore with default parameters (*a=0.5, b=1, c=0.5, d=1, f=0, w=0).* The transformed data was visually inspected to confirm the were no errors (Methods Figure 2). To combine the nine different MBO samples and account for batch effects, the signals were aligned using an unbiased approach, the gaussNorm function in flowStats (Hahne et al., 2013). Local maxima are detected above the bandwidth we set to be above 0.05, to avoid picking up noise, each peak is given a confidence score reflecting the height and sharpness of the peak, the threshold for two peaks to be considered too close together was set too 0.05. Landmarks are then detected and aligned, such that each landmark is shifted to a benchmark, which corresponds to the position of the closest peaks across all samples. After alignment the data is reverse transformed to improve visualization by UMAP in downstream analysis.

6.04 Creation of cell type clusters

For the analysis in Figure 3, to test cluster methods and cell type annotation methods, we selected a subset of hMO cells. From 8 of the MBO 9000 cells were randomly selected and one sample all the cells (1578) cells were selected before transformation and alignment. We compared FlowSom(ref), Phenograph(Ref) and the Seurat(Ref) Louvain network detection function as well as parameter space (k neighbours, resolution, k clusters) available for the different algorithms. We calculated intrinsic statistics and produced UMAPs and heatmaps for visualization. We found FlowSom was not suitable for creating clusters based on a cell types, although the intrinsic statistics are best for FlowSom clustering (Methods Figure 3). Phenograph uses the Louvain network detection method(ref) and computes the Jaccard coefficient which considers the number of common neighbours between cells. Phenograph functions well, however we saw little difference to the Louvain using the Seurat library and proceeded to use the Seurat package for Louvain network detection to obtain clusters for ease of use with the overall workflow. We then proceeded to test the cluster stability at different resolutions, calculating the RAND Index and standard deviation of the number of clusters across 100 iterations of clustering with different random start points. The results informed the choice of cluster numbers to annotate.

6.05 Cluster annotation

Cell type annotation was performed on the subset of 9000 cells using visualization and a correlation assignment model (CAM) we created. For visualization we created functions to make UMAPs for expression levels of each antigen targeted in the antibody panel as well as heatmaps grouped by cluster numbers. The expended expression patterns were of the antibodies as used in combination with the CAM predictions. For the 2D cultures in Figure 2, cell types were assigned by the visualization of expression values, the known original cell type and the overlap in space on the UMAP. In our culture system iPSC can be come any cell type, NPCs are precursor cells for all three other cell types included (astrocytes, DA neurons and oligodendrocytes). The NPC cultures are multipotent but will contain cells that are beginning lineage selection and those retaining a multi potent state.

For the full hMO dataset of nine samples and the followup hMO datasets used for gating and sorting experiments a Random Forest Model trained on the subset hMO data and Seurat transfer labels predictions were used in addition to the CAM and visualization methods used on the subset data. The combined results of methods are more reliable than each method alone. Each of the four methods of annotation are input into the cluster annotate function to automate the cluster annotation process.

6.06 Creation of the predicted expression matrix for antigen proteins in the antibody panel

Astrocytes, oligodendrocyte precursors (OPCs), oligodendrocytes (Oligo), radial glia (RG), endothelial cells, epithelial cells and pericytes are all expected to be present in hMO tissue. Microglia are found in brain tissue but are not expected to be present in MBOs and thus were not included in the reference matrix. In our early tests we found that pericytes were highly overpredicted. We would expect very few if any of these cell types based on previous scRNAseq experiments. Pericytes are not well defined by the FC panel, and we decided to remove these from the reference matrix. We selected expression values for the 13 antigens target by the FC antibody panel. Not all antigens were available from all cell types or databases. Input data was taken from the following public sources: protein expression data scored from the Human Protein Atlas (https://www.proteinatlas.org), bulk RNAseq from human[37], scRNAseq data from human fetal midbrain and other brain tissue from the Human Cell Landscape[38] and cerebral organoids and primary human cells from the UCSC Cell Browser[39]. For the antibody O4, the epitope is a glycoprotein, and the specific corresponding gene is unknown, however the gene NKX6.2 is a marker of mature oligodendrocytes, with expression highly correlated to O4 protein detection.[40] Finally, the FC data acquired in this study from 2D cell cultures, iPSC, neural precursor cells, neurons, astrocytes, and oligodendrocytes. For each data set the values were z-scored then minmax normalized marker by marker to fit between 0 and 1. The mean expression values were calculated separately for scRNAseq organoid data and scRNAseq brain data. Then the mean expression values were then calculated between scRNAseq-hMO, scRNAseq-Brain, RNAseq. Then the mean of that result was calculated with the FC data. The FC data was weighted more highly than the public data sets because it is experimental data collected on protein levels with the exact antibodies used for MBO experiments, however we didn't not generate data on all possible cell types. The predicted expression values were again z-scored then minmax normalized marker by marker to fit between 0 and 1 to be comparable to the transformed FC data to be used in the correlation assignment model.

6.07 Assigning cell type labels to clusters using correlation to the predicted expression matrix

Pearson correlation coefficients, R values were assigned to each cell, correlating the FC intensity expression levels of antibody panel to the predicted expression values in the reference matrix for each cell type expected in the MBO. The R values were calculated for each potential cell type. Then for each hMO cell the max R value and the second max R value were selected. These values were then used to predict the cell type for each hMO cell. A threshold was set of $R > 0.45$ for a cell type to be predicted, otherwise the cell is assigned as 'unknown'. If the $Rmax1 – Rmax2 < 0.05$ then a mix cell type is assigned. For example, Neuron-NPC. To the cluster annotations the top three most frequently predicted cell types for each cluster were calculated. If most of cells were predicted within a cluster as one cell type, this cell type was assigned to the cluster. If the frequencies of predicted cell types were distributed across many cell types, the cluster was assigned as mixed or unknown.

6.08 Random Forest Model

A data frame was created from cell type from the 9000 cells per sample subset of hMO data and the matching expression. The data was split into 50/50 into test and training data. The training data was input into the function RFM_train which uses the *randomForest* and the function caret for optimization. A range of number of variables randomly sampled in each split (mtry) from 1 to 10, the best mtry was 6. Ranges of other parameters were tested, and the optimal values were used in to train the final model: max nodes = 30, node size = 25 and number of trees = 1000. The trained model was then used to predict the cell type of each cell in the full data set and the new flow sorted data. The topmost predicted cell type for each cluster was used as the cluster annotation prediction.

6.09 Label transfer using seurat

We made a function that follows the Seurat workflow for label transfer combined into one function. The annotated Seurat object from the 9000 cells per sample subset of hMO data was used as the reference data and the full dataset and FACS sorted datasets were used as the query objects. Anchors were found between the two objects using 25 principal components to predict the cell types, the max prediction was selected for each cell in the query data. No threshold for predictions were set. The most frequently predicted cell types within each cluster were used as the cluster predictions.

6.10 Quantification of cell types and statistical analysis

Proportionality tests were run using the R library *scProportionTest* (https://github.com/rpolicastro/scProportionTest) using the Seurat object with all annotated cells as the input. One-way ANOVAs, two-way ANOVAs and Tukey's pos hoc tests for main effects and interactions were all run using functions in our R library. A preprocessing function is used to pull the expression data out of the Seurat object and add the desired variables: iPSC line, data of experiment, days in culture, hMO batch. The statistic functions use the base R functions *aov* and *TukeyHSD*. The effect of each variable were analyzed separately. A loop is used to analyze each cell type separately. Two-way ANOVAs were performed with one of the variables listed above and protein (13 targeted in the antibody panel) as second variable.

Methods Table 5: Function names in the CelltypeR library for analysis processes.

Analysis step	**Function name**	**Processes**
Data preprocessing	fsc_to_df	Read in fsc create R list and data frame
	Harmonize	Transform, align, retro-transform, create data frame
	df_to_seurat	Create a Seurat single cell object
Create cell type clusters	Explore_param	Run FlowSom, Phenograph, Louvain (Seurat) with vectors of parameters. Create UMAPs, heatmaps, clustree plot
	Intrinsic_stats	Plot intrinsic stats run in
	Clust_stability	Select one clustering method and compare one parameter. Run cluster stability statistics.
Annotate clusters	See_features	Visualization with UMAP and heatmap
	CAM	Correlation assignment method, predicts cell types, creates plots and tables of the prediction outputs. Requires a reference matrix.
	RFM_train	Input annotated FC dataset to train the Random Forest Model internally optimizing parameters and saving the best model.
	RFM_predict	Predict cell types with RFM, creates plots and tables of the prediction outputs.
	Seurat_transfer	Requires and annotated Seurat object as a reference, creates plots and tables of the prediction outputs.
	Cluster_annotate	Annotates clusters on the consensus of inputs.
Compare Groups	Prep_for_stats	Selects expression data from Seurat object organized by designated variables
	Run_stats	Runs a series of ANOVAs with Tukey's test and outputs results tables
	make_plots	Creates box plots to visualize interaction effects

6.11 Testing gates reverse engineered using hypergate

Cell types were selected in full annotated hMO dataset and input into the *hypergate* function.[28] A table of predictions was output. For each cell type the threshold levels for each antibody required to define the cell type were output. These thresholds are in order from most to least important. For testing the gates, manual gating was applied in FlowJo with the top gate for each cell type in each sample being set as live single cells. The gates were applied in an AIW002 sample and then applied across the other samples. For gating the two antibodies were visualized by scatter plot and a box was drawn selecting the thresholded cells from the antibody pair. The gated cells were then selected and gated with the next pair of antibodies until all thresholds were applied. The final gated cell types from all samples were exported as fsc files and read into R following the CelltypeR workflow. To apply gates to FACS four selected samples examined each cell type gate and selected gates which mostly exclusive for different cell types. The neurons can be separated from glia and then split into two populations and the glia can be split into two populations.

6.12 Single cell sequencing analysis

The FASTQ files processed using 10X CellRanger 5.0.1 software are installed on the Digital Research Alliance of Canada: Beluga computing cluster. For each of the four sorted populations, the CellRanger output files raw expression matrix, barcode, and feature files were used to create a Seurat data object with minimum filtering of RNA features > 100. After this point data was run locally and all details can be found in the R notebook, 'scRNAseq_processing'. RNA features, RNA counts, and percent mitochondria were checked for quality control for each sample: Neurons1, Neurons2, Glia1(astrocytes) and Glia2 (radial glia). Further filters were applied.

Methods Table 6: Filters applied to each population

Sample	RNA features	RNA count	Percent Mitochondria
Neurons1	>250	<10000	<25
Neurons2	>300	<10000	<25
Astrocytes	>500	<10000	<25
Radial Glia	>300	<10000	<25

For the glia samples there was a large number of cells after filtering. The Seurat function HTODemux was used to assign Hashtag (replicate labels). For neuron samples and radial glia all cells were selected, for glia1/astrocyte sample the original count was very high. Increase selection of true cells, cells with assigned hashtags were used for further processing. For all samples, doublets were removed using Doublet Finder (ref https://www.cell.com/cell-systems/fulltext/S2405-4712(19)30073-0).[41] The expected percent of doublets estimation based on the number of cells present after filtering and the 10X version 3 user guide. For each sample data was normalized, variable features selected, PCA and UMAP dimensional reductions were performed, and clusters detected with Louvain network detection (25 dimensions and 43 neighbours selected, and a range of resolutions was run).

Clusters were annotated using a consensus between expression of known cell type markers from gene lists, analysis of cluster markers and cell type predictions of reference data (see below) using Seurat find anchors and label transfer. Subtypes of major cell type groups were observed, at this point these cluster were all merged. The individually processed samples were then merged, samples were down sampled to balance the data and ease processing time.

Methods Table 7: Cell counts after filtering and counts used in merged data object

Sorted Population	Approximate number of cells loaded in 10X Controller	Cell count after filtering	Down sample to
Neurons1	10000	1809	1809
Neurons2	15000	9390	2000
Astrocytes	15000	8123	3000
Radial Glia	15000	4805	2000

After the four samples were merged the standard processing and clustering was run again using the same settings. Clusters were annotated again, retaining subtypes of each cell type and identifying the DA neurons. Each subtype was analyzed to find subtype markers and analyze using GO biological processes. Reference datasets (see below) using Seurat anchors and label transfer predictions were used to define subtypes of cells. A threshold for assignment was set to 0.5 for brain reference data and 0.8 for hMO scRNAseq data.

Methods Table 8: Data sets used as reference data to predict cell types in scRNAseq

Data set	Reference	Link to data
Developing cortex	Nowakowski et al, Science 2017[42]	https://cortex-dev.cells.ucsc.edu
Developing forebrain	Van Bruggen et al, Developmental Cell 2022[43]	https://human-forebraindev.cells.ucsc.edu
Developin brain	Bhaduri et al, Nature 2021[44]	https://dev-brain-regions.cells.ucsc.edu
Adult midbrain DA subtypes Astrocyte subtypes	Kamath et al, Nature 2022[45]	https://singlecell.broadinstitute.org/ single_cell/ study/SCP1768/
Human Midbrain Organoids	Mohamed et al, BrainComm 2021[46]	https://www.ncbi.nlm.nih.gov/geo/query/ acc.cgi?acc=GSE186780

Developing cortex, forebrain and whole brain datasets were all reconstructed into Seurat objects from the UCSC cell browser following the website instructions.[47] Each reference was down sampled in Seurat to reduce the total cell number to less than 50000.

For snRNAseq data from human adult postmortem brains (Kamath et al) three separate reference sets were created. The expression matrix, barcodes and feature files were used to create a Seurat object. The meta for cell type and cell subtype annotations data was added from the UMAP_tsv files provided by Kamath et al. The brain region data was added from the provided meta data file. The adult midbrain was subset by brain region selecting only the midbrain cells. The DA subtypes and astrocyte subtypes were separately subset by using the main cell type annotation.

1. All cell types (astrocytes, oligodendrocytes, microglia, endothelial cells, DA neurons and other neurons). This was used in the initial cell type annotations.
2. DA neuron subtypes, used to try to identify DA subtypes. All the hMO subtypes matched only one subtype from adult brain.
3. Astrocyte subtypes, used to identify astrocyte subtypes. All astrocytes subtypes in hMO matched one subtype.

After annotating the main groups of cell types (DA neurons, neurons, astrocytes, radial glia, NPCs, mixed) subtype annotations were applied. To annotated subtypes, the main cell type was subset. The *Seurat* find all markers function was used allowing both up and down regulated gene markers. The top 5-10 marker genes sorted by highest Log2 Fold change with significant adjusted p-values were further investigated by literature search to determine the cell subtypes.

7.Data availability

Flow cytometry: Raw data and FlowJo selected live gated cells are available on github and deposited at https://flowrepository.org/
scRNAseq: The FASTQ files, CellRanger outputs will be deposited on GEO

8.Code availability

All materials are available on github:
https://github.com/RhalenaThomas/CelltypeR

The repository includes:

1. R library *CelltypeR* containing all functions listed above.
2. Workbooks for each analysis step.
3. Code used to generate figures.

References

1. Mohamed, N.-V. *et al.* Midbrain organoids with an *SNCA* gene triplication model key features of synucleinopathy. *Brain Commun.* **3**, fcab223 (2021).
2. Di Lullo, E. & Kriegstein, A. R. The use of brain organoids to investigate neural development and disease. *Nat. Rev. Neurosci.* **18**, 573 (2017).
3. Wray, S. Modelling neurodegenerative disease using brain organoids. in *Seminars in Cell & Developmental Biology* vol. 111 60–66 (Elsevier, 2021).
4. Fiorenzano, A. *et al.* Single-cell transcriptomics captures features of human midbrain development and dopamine neuron diversity in brain organoids. *Nat. Commun.* **12**, 1–19 (2021).
5. Andrews, T. S. & Hemberg, M. Identifying cell populations with scRNASeq. *Mol. Aspects Med.* **59**, 114–122 (2018).
6. Chen, G., Ning, B. & Shi, T. Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* **10**, 317 (2019).
7. Jiang, R., Sun, T., Song, D. & Li, J. J. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biol.* **23**, 1–24 (2022).
8. Fernández-Zapata, C., Leman, J. K., Priller, J. & Böttcher, C. The use and limitations of single-cell mass cytometry for studying human microglia function. *Brain Pathol.* **30**, 1178–1191 (2020).
9. Nguyen, Q. H., Pervolarakis, N., Nee, K. & Kessenbrock, K. Experimental considerations for single-cell RNA sequencing approaches. *Front. Cell Dev. Biol.* **6**, 108 (2018).
10. Bremond Martin, C., Simon Chane, C., Clouchoux, C. & Histace, A. Recent Trends and Perspectives in Cerebral Organoids Imaging and Analysis. *Front. Neurosci.* **15**, 717 (2021).
11. Albanese, A. *et al.* Multiscale 3D phenotyping of human cerebral organoids. *Sci. Rep.* **10**, 1–17 (2020).
12. Brown, M. & Wittwer, C. Flow cytometry: principles and clinical applications in hematology. *Clin. Chem.* **46**, 1221–1229 (2000).

13. Drouet, M. & Lees, O. Clinical applications of flow cytometry in hematology and immunology. *Biol. Cell* **78**, 73–78 (1993).

14. Woo, J., Baumann, A. & Arguello, V. Recent advancements of flow cytometry: new applications in hematology and oncology. *Expert Rev. Mol. Diagn.* **14**, 67–81 (2014).

15. Pruszak, J., Sonntag, K.-C., Aung, M. H., Sanchez-Pernaute, R. & Isacson, O. Markers and methods for cell sorting of human embryonic stem cell-derived neural cell populations. *Stem Cells* **25**, 2257–2268 (2007).

16. Turaç, G. *et al.* Combined flow cytometric analysis of surface and intracellular antigens reveals surface molecule markers of human neuropoiesis. *PLoS One* **8**, e68519 (2013).

17. Woodard, C. M. *et al.* iPSC-derived dopamine neurons reveal differences between monozygotic twins discordant for Parkinson’s disease. *Cell Rep.* **9**, 1173–1182 (2014).

18. Yuan, S. H. *et al.* Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. *PloS One* **6**, e17540 (2011).

19. Cheung, M. *et al.* Current trends in flow cytometry automated data analysis software. *Cytometry A* **99**, 1007–1021 (2021).

20. Saeys, Y., Van Gassen, S. & Lambrecht, B. N. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat. Rev. Immunol.* **16**, 449–462 (2016).

21. Chen, C. X.-Q. *et al.* A Multistep Workflow to Evaluate Newly Generated iPSCs and Their Ability to Generate Different Cell Types. *Methods Protoc.* **4**, 50 (2021).

22. Mohamed, N.-V. *et al.* Generation of human midbrain organoids from induced pluripotent stem cells. *MNI Open Res.* **3**, 1 (2021).

23. Mohamed, N.-V. *et al.* Microfabricated disk technology: Rapid scale up in midbrain organoid generation. *Methods* **203**, 465–477 (2022).

24. Monzel, A. S. *et al.* Derivation of human midbrain-specific organoids from neuroepithelial stem cells. *Stem Cell Rep.* **8**, 1144–1154 (2017).

25. Atamian, A., Cordón-Barris, L. & Quadrato, G. Taming human brain organoids one cell at a time. in *Seminars in Cell & Developmental Biology* vol. 111 23–31 (Elsevier, 2021).

26. Xiang: hESC-derived thalamic organoids form reciprocal... - Google Scholar. https://scholar.google.com/scholar_lookup?title=hESC-Derived%20Thalamic%20Organoids%20Form%20Reciprocal%20Projections%20When%20Fused%20with%20Cortical%20Organoids&publication_year=2019&author=Y.%20Xiang&author=Y.%20Tanaka&author=B.%20Cakir&author=B.%20Patterson&author=K.Y.%20Kim&author=P.%20Sun&author=Y.J.%20Kang&author=M.%20Zhong&author=X.%20Liu&author=P.%20Patra.

27. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).

28. Becht, E. *et al.* Reverse-engineering flow-cytometry gating strategies for phenotypic labelling and high-performance cell sorting. *Bioinformatics* **35**, 301–308 (2019).

29. Yoo, J.-E. *et al.* Trophoblast glycoprotein is a marker for efficient sorting of ventral mesencephalic dopaminergic precursors derived from human pluripotent stem cells. *Npj Park. Dis.* **7**, 1–11 (2021).

30. Chen, X., Rocha, C., Rao, T. & Durcan, T. NeuroEDDU protocols_iPSC culture_interactive protocol. (2019) doi:10.5281/zenodo.3733913.

31. Chen, X., Lauinger, N., Rocha, C., Rao, T. & Durcan, T. M. Generation of dopaminergic or cortical neurons from neuronal progenitors (interactive protocol). (2019) doi:10.5281/zenodo.3733914.

32. Jefri, M. *et al.* Stimulation of L-type calcium channels increases tyrosine hydroxylase and dopamine in ventral midbrain cells induced from somatic cells. *Stem Cells Transl. Med.* **9**, 697–712 (2020).

33. Piscopo, V. & Durcan, T. M. Generation of OPCs and Oligodendrocytes from iPSCs. (2020) doi:10.5281/zenodo.3745597.

34. Douvaras, P. & Fossati, V. Generation and isolation of oligodendrocyte progenitor cells from human pluripotent stem cells. *Nat. Protoc.* **10**, 1143–1154 (2015).

35. Soubannier, V. *et al.* Rapid Generation of Ventral Spinal Cord-like Astrocytes from Human iPSCs for Modeling Non-Cell Autonomous Mechanisms of Lower Motor Neuron Disease. *Cells* **11**, 399 (2022).

36. Mohamed, N.-V. *et al.* Generation of human midbrain organoids from induced pluripotent stem cells. *MNI Open Res.* **3**, 1 (2019).

37. 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).

38. Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).

39. Bhaduri, A. *et al.* Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).

40. Elbaz, B. & Popko, B. Molecular control of oligodendrocyte development. *Trends Neurosci.* **42**, 263–277 (2019).

41. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).

42. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

43. van Bruggen, D. *et al.* Developmental landscape of human forebrain at a single-cell level identifies early waves of oligodendrogenesis. *Dev. Cell* **57**, 1421–1436 (2022).

44. Bhaduri, A. *et al.* An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).

45. Kamath, T. *et al.* Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson’s disease. *Nat. Neurosci.* **25**, 588–595 (2022).

46. Mohamed, N.-V. *et al.* Midbrain organoids with an SNCA gene triplication model key features of synucleinopathy. *Brain Commun.* **3**, fcab223 (2021).

47. Speir, M. L. *et al.* UCSC Cell Browser: visualize your single-cell data. *Bioinformatics* **37**, 4578–4580 (2021).

48. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl. Acad. Sci.* **97**, 14720–14725 (2000).

49. Pruszak, J., Ludwig, W., Blak, A., Alavian, K. & Isacson, O. CD15, CD24, and CD29 define a surface biomarker code for neural lineage differentiation of stem cells. *Stem Cells* **27**, 2928–2940 (2009).

50. Sundberg, M. *et al.* CD marker expression profiles of human embryonic stem cells and their neural derivatives, determined using flow-cytometric analysis, reveal a novel CD marker for exclusion of pluripotent stem cells. *Stem Cell Res.* **2**, 113–124 (2009).

51. Wang, J., O'Bara, M. A., Pol, S. U. & Sim, F. J. CD133/CD140a-based isolation of distinct human multipotent neural progenitor cells and oligodendrocyte progenitor cells. *Stem Cells Dev.* **22**, 2121–2131 (2013).

52. Sandor, C. *et al.* Transcriptomic profiling of purified patient-derived dopamine neurons identifies convergent perturbations and therapeutics for Parkinson's disease. *Hum. Mol. Genet.* **26**, 552–566 (2017).

53. Liu, Y. *et al.* CD44 expression identifies astrocyte-restricted precursor cells. *Dev. Biol.* **276**, 31–46 (2004).

54. Jurga, A. M., Paleczna, M., Kadluczka, J. & Kuter, K. Z. Beyond the GFAP-astrocyte protein markers in the brain. *Biomolecules* **11**, 1361 (2021).

55. Henrik Heiland, D. *et al.* Tumor-associated reactive astrocytes aid the evolution of immunosuppressive environment in glioblastoma. *Nat. Commun.* **10**, 1–12 (2019).

56. McPhie, D. L. *et al.* Oligodendrocyte differentiation of induced pluripotent stem cells derived from subjects with schizophrenias implicate abnormalities in development. *Transl. Psychiatry* **8**, 1–10 (2018).

Figure 1: A workflow to identify and quantify cell types in midbrain organoids (hMOs) using a panel of flow cytometry antibodies. A) Schematic of the CelltypeR workflow: tissue (hMO) is dissociated and labelled with an antibody panel, expression levels are measured on individual cells using flow cdytometry (FC), live single cells are gated from the debris and doublets in FlowJO. The data is then preprocessed in R, merging files, and harmonizing the data if wanted. Unsupervised clustering is used to find groups of cell types, methods are provided to aid in cluster annotation, annotated cells are quantified, and statistical analysis is applied. B) Example image of a cryosection from an AJG001C hMO, 285 days in final differentiation culture, showing total nuclei (Hoechst), oligodendrocytes (O4), astrocytes (GFAP) and neurons (MAP2). Top: cross section of a whole hMO stitched together from tiled images, scale bar = 250mm. Bottom: zoomed in image cropped from the whole hMO image, scale bar = 250mm. C) Density plots showing the cell size on the y-axis (FSC) and intensity of staining for each antibody in the panel, x-axis (log scale biexponential transformation).

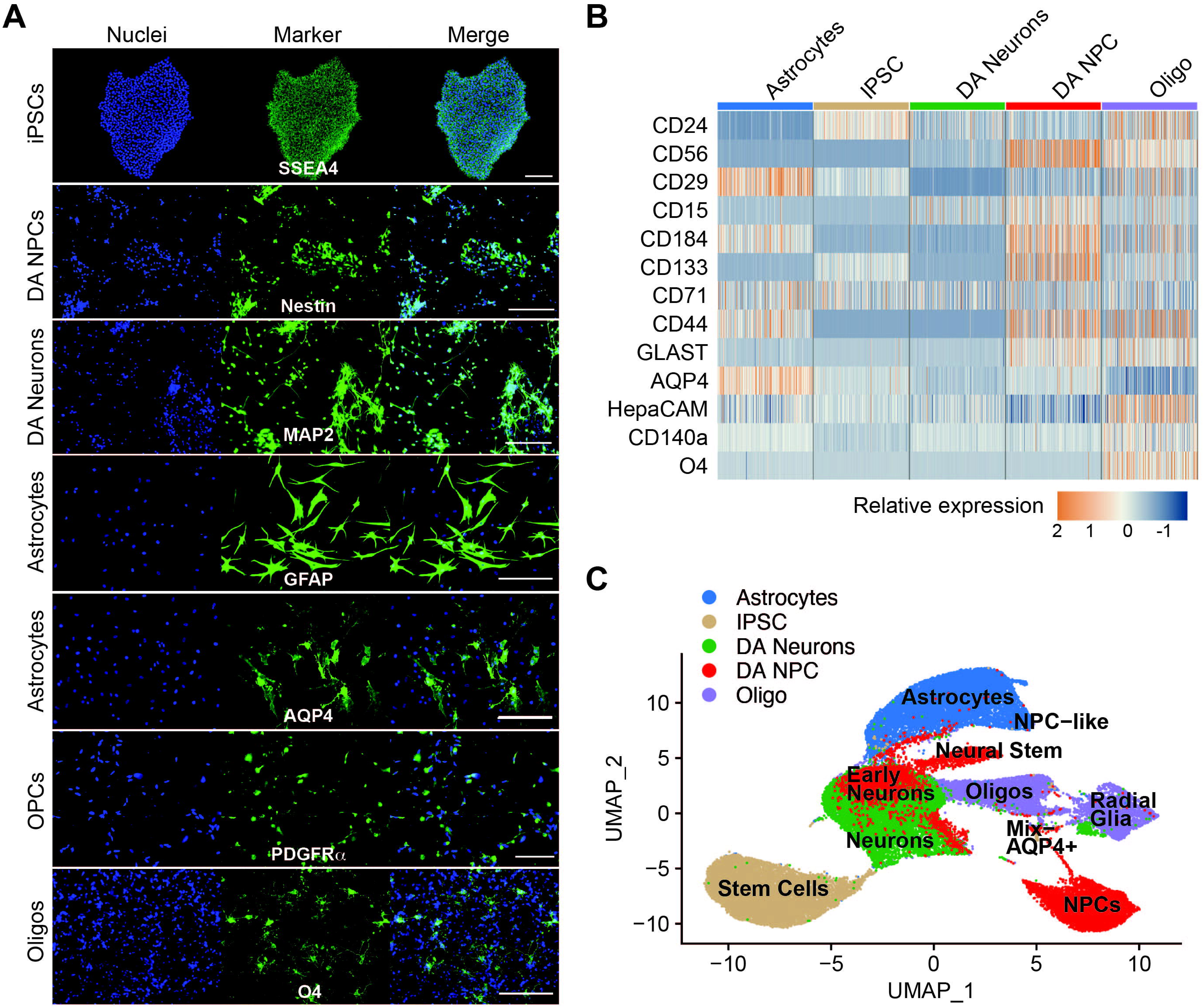

Figure 2: The antibody panel can be used to identify cell types expected to be present in hMOs. A) Example images of different brain cell types (indicated on the left) derived from the healthy control AIW002-02 iPSC line and individually differentiated. Cell cultures were stained with a cell type specific marker (green) and Hoechst (blue) for nuclei. Scale bars 200mM. B) Heatmap of the normalized intensity of protein levels measured by FC for a subset of cells from each cell culture (indicated above), antibodies used to measure protein levels are indicated on the left. C) UMAP of merged cell cultures (indicated by colour and in the legend) showing the separation and overlap of cell types. Louvain network detection was applied, and clusters were annotated based on protein marker expression, the cell type annotations are indicated by labels on the UMAP.

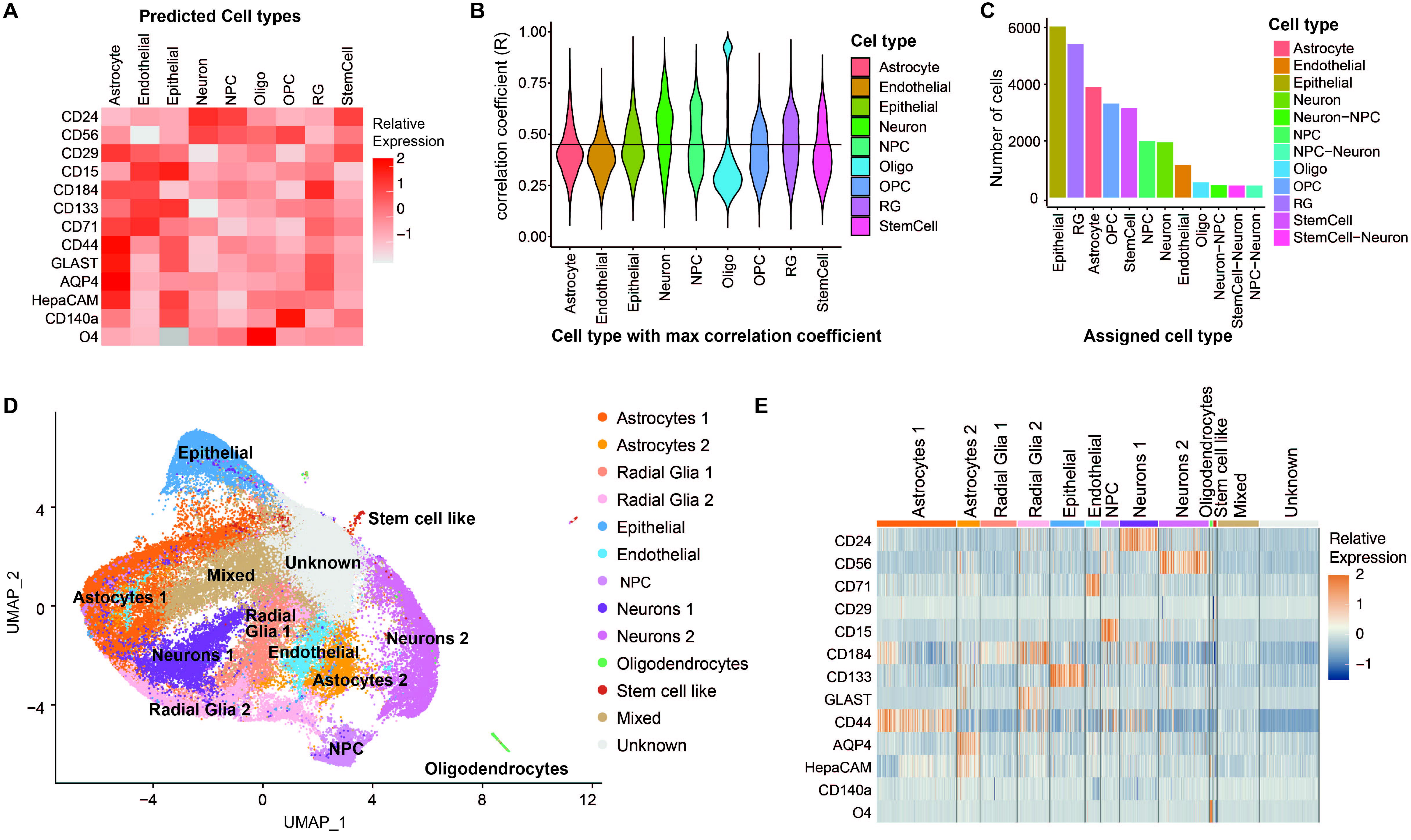

Figure 3: Identification of cell types in hMO using the flow cytometry antibody panel. A) Heatmap of predicted relative expression of each antibody in the FC panel for each potential cell type in hMOs. Values are calculated from 2D FC intensities, scRNAseq from hMOs and human brain and RNAseq from human brain. B) Violin plot showing the distribution Pearson's correlation coefficients R for hMO cells (y-axis) with the indicated potential brain cell type (x-axis). The R values are plotted for the cell type with the max R value. The black line indicates the threshold of R=0.45 which was set as the cut-off for assigning a cell type prediction. C) Bar chart showing the number of hMO cells categorized as each cell type by the max correlation, each cell type is indicated on the x-axis. HMO cells were assigned as a double cell type if the first and second max R values were within 0.05. Only cell assignments with over 100 cells are included in the bar chart. D) UMAP showing unsupervised clustering by Louvain network detection. Cell types were annotated using a combination of correlation assignment and expert analysis of expression within clusters. E) Heatmap of relative expression of each antibody grouped by the cell types identified by unsupervised clustering of hMO cells. (n=73,578 cells from 9 hMO samples).

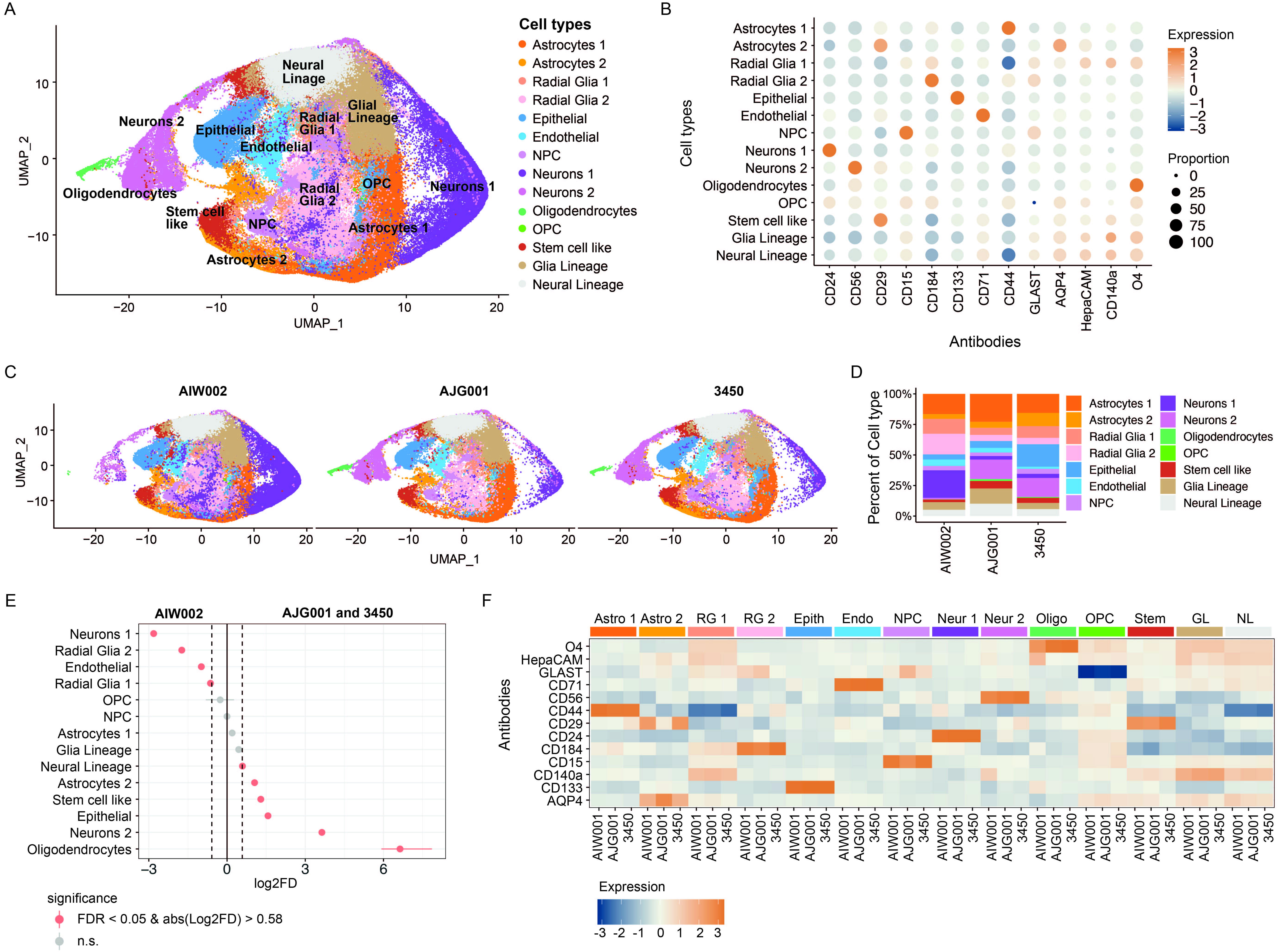

Figure 4: Differences in cell types and protein expression between three healthy control donor iPSCs. A) UMAP of the full dataset from nine hMO samples (n > 197,000 cells) annotated using CelltypeR. B) Dot plot of the expression level (colour intensity) and the proportion of cell (dot size) for each protein marker detected with the panel in each cell type group. C) UMAP split by iPSC line (3 samples pooled per iPSC line) showing the proportion of cells in each iPSC line. Cell annotations and colours are the same as the UMAP above from A. D) Bar chart of the proportion of hMO cells in each cell type (indicated by colour) for each iPSC line (x axis). Colours corresponding to cell types are shown in the legend on the right. E) Dot plot with confidence interval for the proportionality test comparing the AIW002 iPSC line to the AJG001 and 3450 iPSC lines, for each cell type (y-axis). Pink dots indicate a significant difference in cell type proportion (FDR < 0.05 and absolute value of Log2FD > 0.58). Negative log2FD values indicate cell proportions increased in AIW002 and positive values indicate cell proportions decreased in AIW002 compared to the other two iPSC lines. F) Heatmap of mean protein expression values grouped by cell type and split into the three iPSC lines. Line names are indicated on the bottom x-axis and cell types are indicated on the top x-axis.

Figure 5: CelltypeR can be used to gate populations of interest and annotate the gated cells. A) Schematic showing the method used to gate cell type populations defined with CelltypeR. Cell types were annotated and selected in the full hMO dataset. Then the package hypergate was applied to reverse engineer the threshold expression levels to define each cell population. Astrocytes, radial glia, stem cell like cells, oligodendrocytes, epithelia cells, endothelial cells, neural precursor cells (NPCs), neurons 1 and neurons 2 cell populations were separately gated manually in FlowJo from the hMO dataset. B) The proportions of FlowJo gated cells is uneven across populations, to improve visualization, the gated populations were down sampled to 10 000 cells per cell type. Some cell types have fewer cells and the total population was included. The UMAP of the merged and clustered FlowJo gated cells is coloured by the gated populations. Labels on the UMAP are the cell types annotated using the CelltypeR workflow. C) Bar chart with the proportion of cell types identified with CelltypeR (indicated by colour in the legend) within each FlowJo gated population (x-axis).

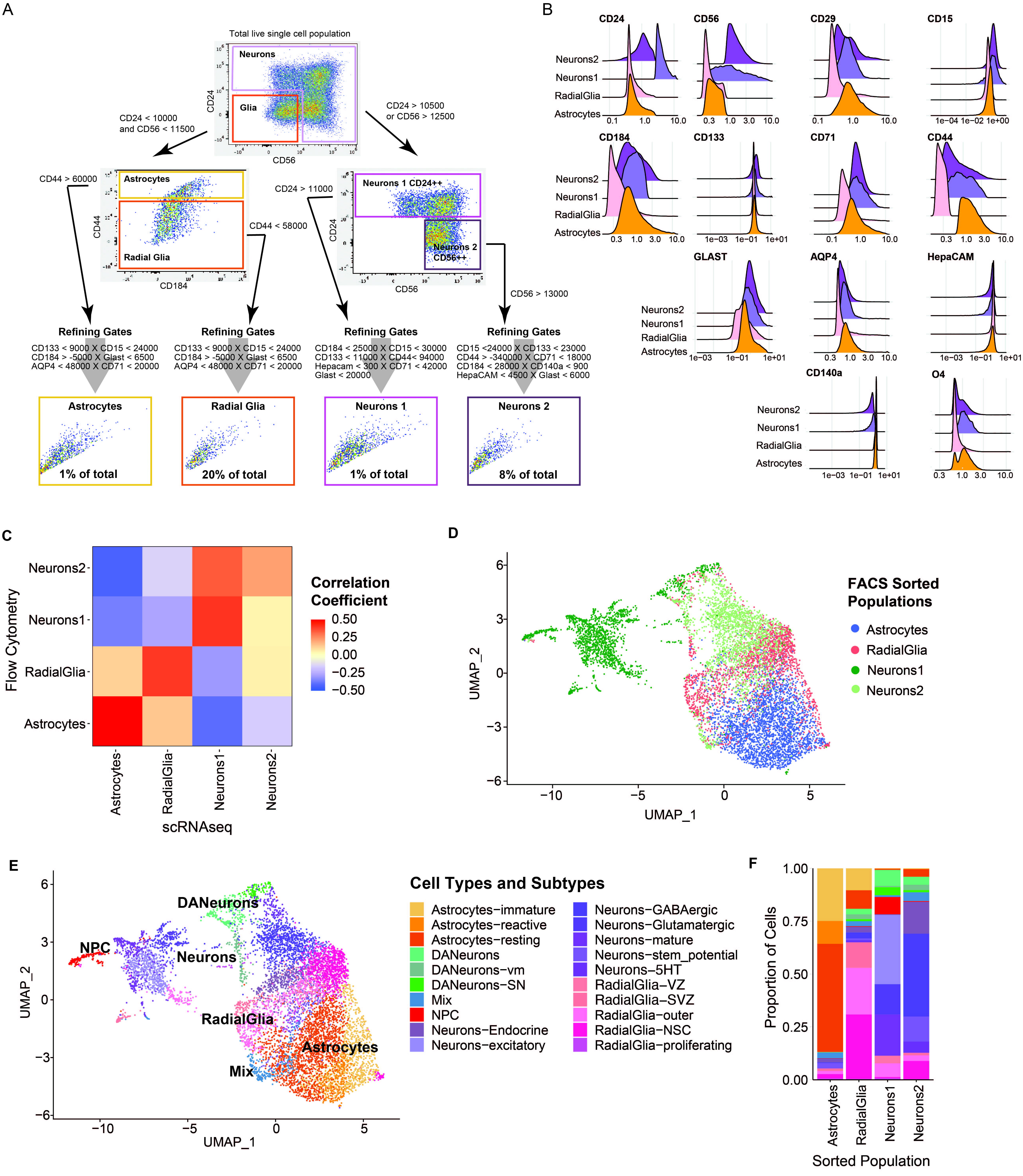

Figure 6: scRNAseq analysis of four FACS sorted populations defined using CelltypeR confirms cell types and provides transcriptional profiles for these cell populations. A) FlowJo gating strategy applied to new hMO derived cells to isolate four cell populations by FACS: neurons1, neurons2, astrocytes and radial glia. The approximate proportion of cells gated in each final sorted population is indicated in the gating box. B) Protein expression levels measured by FC antibody intensity for each FACS gated cell population. C) Correlation of RNA transcript expression of genes corresponding to the 13 protein markers used for FACS sorting. Note there is a high correlation between RNA expression and protein expression for radial glia, neurons 1 and astrocytes. Only the astrocytes correlation has a statistically significant correlation. The neurons2 protein expression correlates more strongly with the neurons1 RNA expression. D) UMAP of the four sorted populations merged and clustered with Louvain network detection. Neurons1 has only 1809 cells, neurons2 was down sampled to 2000, astrocytes were down sampled to 3000 and radial glia were down sampled to 2000 to improve visualization. The original FAC sorted population is indicated by colour and in the legend. E) UMAP of the four merged populations with cell types and cell subtypes annotated from the scRNAseq data. The UMAP is coloured by cell subtypes and the main cell types are labelled on top of the UMAP. Subtype markers were identified from differential RNA between clusters. F) Bar chart of the proportion of cell subtypes annotated from RNA expression within each FACS sorted populations, cell subtypes are coloured by the same legend as the UMAP in E.