Differential growth and transcriptomic profile of stem cell-derived midbrain astrocytes

Zongze Li[1&2], Lucia Fernandez Cardo[1], Michal Rokicki[1], Jimena Monzón-Sandoval[1], Viola Volpato[1], Frank Wessely[1], Caleb Webber[1*], Meng 89Li[2*]

[1] Dementia Research Institute, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ, UK

[2] Neuroscience and Mental Health Innovation Institute, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ, UK

*Corresponding authors:

Meng Li, lim26@cardiff.ac.uk,

Caleb Webber, webberc4@cardiff.ac.uk

Key words: astrocyte, regional heterogeneity, ventral midbrain, dopaminergic, single cell RNA-sequencing

1

16   **Abstract**

17   Regional specificity of stem cell-derived astrocytes is believed to be an important

18   prerequisite for their applications in disease modelling and cell-based therapies. The

19   regional identity of these astrocytes is often defined by the positional characteristics

20   of their antecedent, stem cell-derived neural progenitors patterned to a fate of

21   interest, with the assumption that the positional specification is to be preserved by

22   the derived astrocytes. Using a human induced pluripotent stem cell line designed

23   for tracing midbrain floor plate derivatives, here we show that lineage composition of

24   the derived astrocytes is not a faithful recapitulation of the founder progenitor

25   population, as demonstrated by the loss of floor plate differentiated progeny in the

26   final astrocyte products. Using deep single cell RNA sequencing, we identified

27   distinct transcriptomic signatures of midbrain floor plate-derived astrocytes. Our

28   study highlights the need for rigorous characterisation of pluripotent stem cell-

29   derived regional astrocytes and provides a valuable resource for assessing midbrain

30   floor plate-derived human astrocytes.

Introduction

31

32 Astrocytes are the most abundant cells in the brain. They play important roles in the

33 central nervous system in supporting neuronal survival and synaptic activities,

34 including the regulation of ionic homeostasis, providing energetic support, elimination

35 of oxidative stress, and neurotransmitter removal and recycle (Verkhratsky and

36 Nedergaard 2018). Abnormalities in astrocytes have been linked to various

37 neurodegenerative and neurodevelopmental disorders, such as Parkinson's disease,

38 Alzheimer's disease, Huntington's disease, autism spectrum disorders and

39 Alexander's disease (Molofsky et al. 2012; Phatnani and Maniatis 2015; Booth et al.

40 2017). There is therefore a growing interest in using human pluripotent stem cell

41 (PSC)-derived astrocytes for disease modelling *in vitro* (Chandrasekaran et al. 2016).

42 Contrary to the widely held belief that astrocytes in the brain are largely identical,

43 recent studies revealed diversity in their transcriptomic profile, physiological properties

44 and function (Oberheim et al. 2009; Schober et al. 2022). For example, single cell and

45 spatial transcriptomic studies have identified several astrocyte subpopulations in the

46 mouse cortex (Zhu et al. 2018; Batiuk et al. 2020; Bayraktar et al. 2020). In human,

47 although astrocyte heterogeneity remains largely elusive, heterogeneity in radial glia

48 across brain regions and within midbrain has been reported (La Manno et al. 2016; Li

49 et al. 2023). Furthermore, different molecular and physiological features and distinct

50 responses to stimuli were also observed in astrocytes from different mouse brain

51 regions (Takata and Hirase 2008; Chai et al. 2017; Morel et al. 2017; Itoh et al. 2018;

52 Kostuk et al. 2019; Makarava et al. 2019; Xin et al. 2019; Lozzi et al. 2020). Indeed,

53 astrocyte heterogeneity has been suggested to underly the regional susceptibility to

54 human diseases (Schober et al. 2022). Therefore, recapitulating astrocyte regional

55 specificity in PSC-derived astrocytes is generally accepted as an important

56 prerequisite.

57 Several reports have described the generation of regional astrocytes from human

58 embryonic stem cells or induced pluripotent stem cells (iPSCs) which include forebrain

59 (Krencik et al. 2011; Zhou et al. 2016; Tcw et al. 2017; Lin et al. 2018; Bradley et al.

60 2019; Hedegaard et al. 2020; Peteri et al. 2021), ventral midbrain (Booth et al. 2019;

61 Barbuti et al. 2020; Crompton et al. 2021; de Rus Jacquet et al. 2021), hindbrain and

62 spinal cord (Roybon et al. 2013; Serio et al. 2013; Holmqvist et al. 2015; Bradley et al.
63 2019; di Domenico et al. 2019; Yun et al. 2019). The regional identity of these
64 astrocytes is typically evaluated at the stage of early neural progenitors generated via
65 cell type- or region-directed neural patterning protocols, with the assumption that the
66 positional characteristics will be faithfully preserved in the final astrocyte products.
67 However, astrocyte production in vitro involves an extended period of astrocytic fate
68 induction and progenitor expansion using FGF and EGF while substantial literature
69 reported alterations in region-specific gene expression and/or neurogenic competence
70 in expanded neural progenitors (Jain et al. 2003; Sun et al. 2008; Koch et al. 2009;
71 Falk et al. 2012). Therefore, better characterisations of PSC-derived astrocytes and
72 their lineage-specific features are needed to advance our knowledge about the
73 molecular heterogeneity of human astrocytes.

74 Using a human iPSC line that allows the tracing of LMX1A expressing midbrain floor
75 plate neural progenitors and their differentiated progeny (Cardo et al. 2023), we
76 discovered an unexpected gradual depletion of LMX1A$^+$ progenitor progeny during
77 astrocyte induction from a bulk population of ventral midbrain patterned progenitors
78 despite LMX1A$^+$ progenitors being the predominant starting population. LMX1A$^+$
79 progenitor derived astrocytes can however be generated if astrocytic induction is
80 initiated from purified LMX1A$^+$ progenitors, indicating that the positional constituents
81 of the founding cell population may not be preserved faithfully in the derived
82 astrocytes. Single cell RNA sequencing (scRNAseq) of astrocytes derived from both
83 parental populations identified distinct transcriptomic signatures, providing a useful
84 resource for assessment of human PSC-derived midbrain astrocytes.

Results

Depletion of LMX1A$^+$ progenitors and/or derivatives in ventral midbrain patterned neural progenitor cultures during astrogenic induction

89 To investigate whether regionally patterned neural progenitors retain their lineage
90 identity during astrogenic induction and glial progenitor expansion, we made use of

4

91   the LMX1A-Cre/AAVS1-BFP iPSCs tracer line, which enables the tracking of LMX1A$^+$

92   midbrain floor plate progenitors and their differentiated progeny(Cardo et al. 2023)꘎.

93   We differentiated the LMX1A-Cre/AAVS1-BFP iPSCs towards the ventral midbrain

94   fate following a modified protocol based on Jaeger *et al.* and Nolbrant *et al.*(Jaeger et

95   al. 2011; Nolbrant et al. 2017)꘎(Figure 1A). Immunocytochemistry of day (d)19

96   cultures confirmed a high proportion of cells expressing BFP (96.01±0.42%) and

97   ventral midbrain progenitor markers LMX1A (92.94±0.91%), FOXA2 (94.76±0.57%)

98   and OTX2 (97.82±0.28%; Figure 1B-D with the original images shown in Figure S1A).

99   Most cells (91.26±1.64%) co-expressed LMX1A and FOXA2 (Figure 1B and 1D). At

100   this stage, all BFP$^+$ cells also stained positive to a pan neural progenitor marker

101   NESTIN (Figure 1C). We detected a small proportion of cells expressing midbrain

102   basal plate marker NKX6.1 (2.03±0.47%, Figure 1B-C) whose expression domain in

103   the early developing ventral midbrain partially overlap with that of (Andersson et al.

104   2006)(Andersson et al. 2006)꘎. However, few PAX6$^+$ cells were present (Figure 1C),

105   which marks the forebrain and (Duan et al. 2013)(Duan et al. 2013)꘎.

106   Immunocytochemical analysis of FACS-sorted BFP$^+$ cells confirmed highly enriched

107   expression of LMX1A (95.51±0.09%) and FOXA2 (95.35±0.14%), and co-expression

108   of both markers (94.68±0.10%; Figure 1B and 1D). In contrast, only a small number

109   of LMX1A$^+$ cells (4.96±0.70%) were present in the sorted BFP$^-$ population (Figure 1B

110   and1D). These findings provide further support that BFP expression faithfully

111   recapitulate LMX1A expression and that LMX1A$^+$ ventral midbrain progenitors

112   represent the major cell population in d19(Cardo et al. 2023)(Cardo et al. 2023)꘎.

113   The d19 cells were then induced to undergo astrogenic switch in media containing

114   FGF2 and EGF with the BFP content monitored by flow cytometry at each weekly

115   passaging (Figure 1A, representative gating strategy is shown in Figure S1B-E). We

116   found, unexpectedly, a dramatic decrease of BFP$^+$ cell proportion from the starting

117   point of 88.20±1.10% to only 27.59±8.28% at d43 and to nearly absent as the

118   differentiation continued (Figure 1E). We did not observe evident cell death during

119   culture and replating; thus, the absence of BFP could be either due to the silencing of

120   BFP expression in the derivatives of LMX1A$^+$ progenitors or the loss of these cells

121   through growth competition. To address this question, we performed progenitor

122   expansion and astrogenic induction under the same culture condition with purified d19

5

123   BFP$^+$ progenitors isolated using fluorescence-activated cell sorting (FACS).

124   Interestingly, the sorted BFP$^+$ cells exhibited similar population growth rate to that of

125   unsorted cultures, and the proportion of BFP$^+$ cells remained at approximately 90%

126   throughout the astrogenic induction and glial progenitor expansion period (Figure 1E).

127   This observation demonstrates that BFP expression can be maintained in the

128   derivatives of LMX1A$^+$ midbrain floorplate progenitors, and that the loss of BFP$^+$ cells

129   in the unsorted culture is likely due to their growth disadvantage compared to the

130   derivatives of LMX1A$^-$ progenitors.

131   Our finding is unexpected and demonstrates that the regional or lineage identity of

132   PSC-derived astrocytic cells should not be assumed merely based on the dominant

133   regional property of their cellular origin, given that no *in vitro* fate specification

134   paradigm is 100% efficient.

6
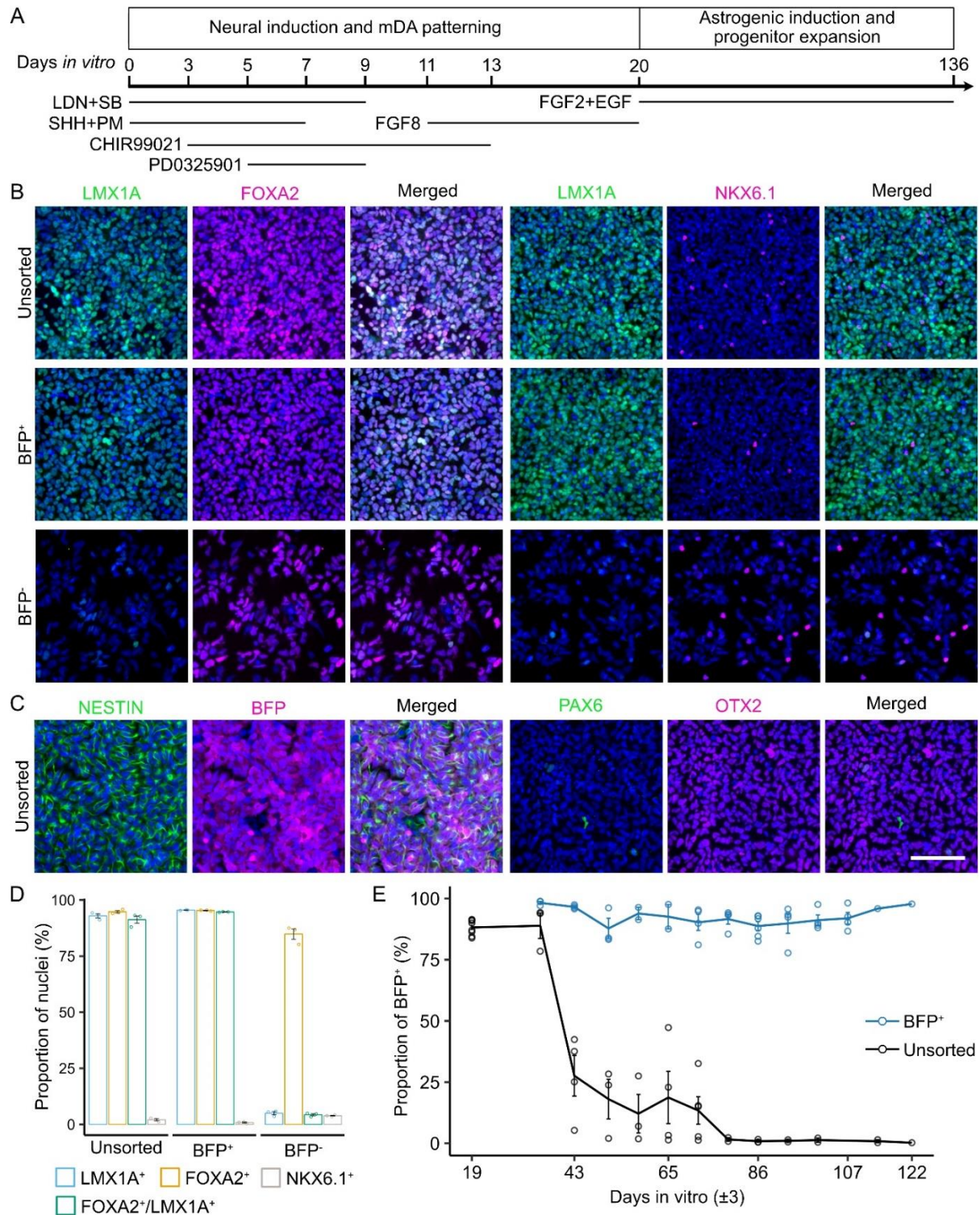
Figure 1. Depletion of LMX1A⁺ progenitors and their derivatives during astrogenic induction in ventral midbrain patterned neural progenitor cultures.

A, Schematic diagram of ventral midbrain neural differentiation and astrogenic induction. B-C, Representative view of immunocytochemistry of ventral midbrain neural progenitor markers and other regional markers in d19 unsorted, sorted BFP⁺, and sorted BFP⁻ population. Scale bar represents 100 μm. Images shown were cropped to 300 μm×300 μm by randomly selecting the region of interest in the nuclei-

7

143 only channel (uncropped greyscale images are shown in Figure S1A). D,
144 Quantification of marker expression in unsorted, sorted BFP$^+$, and sorted BFP$^-$
145 population. Error bars represent the standard error of means (SEM) of three
146 independent experiments. E, Flow cytometry quantification of unsorted and BFP+
147 population during astrogenic induction and progenitor expansion. Each data point
148 represents one biological replicate. The gating strategy used is shown in Figure S1B.

149

Astrogenic switch occurred earlier in derivatives of LMX1A$^+$ midbrain progenitors.

152 Since sorted BFP$^+$ (LMX1A$^+$ or their derivatives) and unsorted astrocytic cultures differ
153 distinctively in BFP expression soon after the initiation of astrogenic induction, for
154 simplicity, these cultures are hereafter referred to as the BFP$^+$ and BFP$^-$ cultures,
155 respectively. Wondering whether the two cell populations behave differently in the
156 process of astrogenic switch, we examined astrocytic marker expression in these
157 cultures at d45 and d98 by immunocytochemistry. SOX9 and NFIA are transcription
158 factors crucial for the initiation of astrogenesis and acquisition of astrogenic
159 competence in the developing central nervous system (Stolt et al. 2003; Deneen et al.
160 2006), while CD44 identifies astrocyte-restricted precursors (Liu et al. 2004). We found
161 that all these markers were more abundantly detected in the BFP$^+$ cultures (NFIA:
162 65.89±2.81%; SOX9: 57.19±4.25%) than the BFP$^-$ cultures (NFIA: 4.26±1.28%; SOX9:
163 8.88±1.82%) at d45 (Two-way ANOVA with post-hoc Turkey test, NFIA: $p=2.52\times10^{-5}$,
164 SOX9: $p=9.21\times10^{-6}$; Figure 2A-B with the original images shown in Figure S2).
165 Although the number of NFIA$^+$ and SOX9$^+$ cells significantly increased in the BFP$^-$
166 cultures by d98 (NFIA: 44.07±4.56% on d98, $p=3.58\times10^{-4}$; SOX9: 44.28±2.84% on
167 d98, $p=1.21\times10^{-4}$), the BFP$^+$ cultures still contained more cells expressing NFIA
168 (65.71±4.25%; $p=2.06\times10^{-2}$) and SOX9 (73.25±2.12%; $p=5.51\times10^{-4}$) than BFP$^-$
169 cultures (Figure 2A and 2C).

170 To investigate whether the temporal difference in astrocytic switch between the BFP$^+$
171 and BFP$^-$ cultures affects maturation and functionality of the derived astrocytes, we
172 initiated astrocyte terminal differentiation by exposing the BFP$^+$ and BFP$^-$ astrocyte
173 precursors to CNTF and BMP4 (Krencik et al. 2011; Bradley et al. 2019) from d87
174 (referred to as early astrocytes) and d136 (late astrocytes). Both BFP$^+$ and BFP$^-$

8

175 cultures exhibited a similar expression profile of classic astrocyte markers, including

176 AQP4, EAAT2, and S100B, but few GFAP$^+$ cells at both time points (Figure 2B and

177 2D with the original images shown in Figure S3).

178 As a reference, we also generated neural progenitors without employing any

179 patterning cues and induced astrogenic switch and astrocyte differentiation from these

180 non-patterned neural progenitors (Figure S4A-B). We found that, while the astrocyte

181 cultures derived from the non-patterned progenitors contained a similar proportion of

182 cells expressing AQP4, EAAT2 and S100B compared to the BFP$^+$ and BFP$^-$ astrocyte

183 cultures, there are more GFAP$^+$ cells in the non-patterned astrocyte preparations

184 (17.89±5.4%, Figure S4A-B).

185 Functional astrocytes exhibit transient calcium (Ca^{2+}) spikes upon chemical

186 stimulation, such as ATP (Zhang et al. 2016). Using a FLIPR Ca^{2+} release assay, we

187 observed a sharp increase in the intracellular Ca^{2+} concentration upon ATP

188 administration in both the early and late BFP$^+$ astrocyte populations (early BFP$^+$:

189 $p=5.37\times10^{-12}$; late BFP$^+$: $p<2.2\times10^{-16}$; Figure 2E-F). ATP induced Ca2$^+$ release is

190 partially mediated by inositol trisphosphate. Indeed, addition of an inositol

191 trisphosphate receptor antagonist 2-aminoethoxydiphenylborate (2-APB) reduced the

192 amplitude (early BFP$^+$: $p=5.88\times10^{-5}$; late BFP$^+$: $p=9.22\times10^{-10}$; Figure 2F) and rise time

193 (early BFP$^+$: $p=2.63\times10^{-5}$; late BFP$^+$: $p=3.59\times10^{-4}$; late BFP$^-$: $p=0.028$; Figure 2G) of

194 ATP-induced Ca2$^+$ response in both the early and late BFP$^+$ astrocytes. Interestingly,

195 the early BFP$^+$ astrocytes had significantly lower peak amplitude than that observed

196 in late BFP$^+$ astrocytes ($p=6.92\times10^{-9}$; Figure 2F), despite their similar level of astrocyte

197 marker expression, suggesting a difference in maturity at the functional level. Early

198 and late BFP$^-$ astrocytes exhibited a similar profile of time-dependent increase in the

199 amplitude of ATP-induced Ca^{2+} response but did not reach statistical significance

200 (Figure 2E-F). However, late BFP$^-$ astrocytes showed a significantly lower peak

201 amplitude than late BFP$^+$ astrocytes ($p<2.2\times10^{-16}$; Figure 2E-F).

202 Taken together, our data demonstrate that astrogenesis occurred earlier in BFP$^+$

203 cultures than BFP$^-$ cells. This temporal difference is also reflected in functional maturity

204 of the derived astrocytes, despite a similar expression profile of classic astroglial

205 markers.
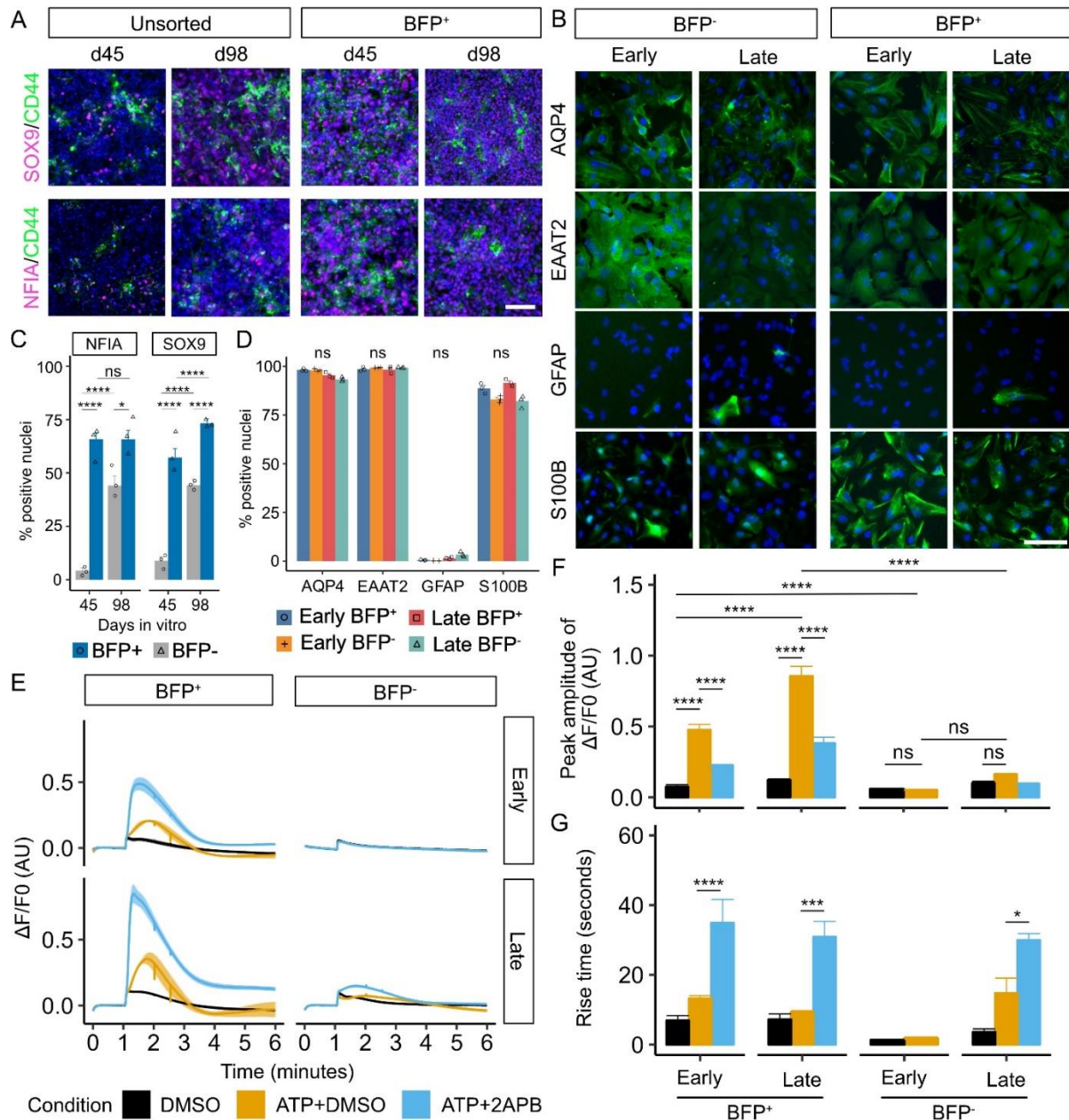
9

Figure 2. Early astrogenic switch and astrocyte maturation in derivatives of LMX1A+ midbrain progenitors.

A, Representative view of immunocytochemistry of astrogenic marker expression in BFP+ and unsorted progenitors at day 45 and 98. Scale bar represents 100 µm. Images shown were cropped to 462 µm × 462 µm by randomly selecting the region of interest in the nuclei-only channel (uncropped greyscale images are shown in Figure S2). B, Representative view of immunocytochemistry of astrocyte marker expression in early and late, BFP+ and BFP- (unsorted) astrocytes. Scale bar represents 100 µm. Images shown were cropped to 300 µm × 300 µm by randomly selecting the region of interest in the nuclei-only channel (uncropped greyscale images are shown in Figure S3). C, Quantification of immunocytochemistry of astrogenic marker in BFP+ and BFP- progenitors at day 45 and 98 shown in Panel A. Error bars represent the standard error of means (SEM) of three independent experiments. Two-way ANOVA was

10

220 performed to compare between lineages (NFIA: p=5.389×10^{-6}, df=1, effect size=3.62;
221 SOX9: p=1.96×10^{-6}, df=1, effect size=4.77) and days of differentiation (NFIA:
222 p=7.82×10^{-5}, df=1, effect size=1.99; SOX9: p=2.62×10^{-5}, df=1, effect size=2.99) D,
223 Quantification of immunocytochemistry of astrocyte marker expression in astrocytes.
224 Error bars represent SEM of three independent experiments. Kruskal-Wallis test
225 results following Bonferroni correction are shown on the top of the figure (AQP4:
226 p.adjust=0.12, df=3, H=8.95; EAAT2: p.adjust=1.00, df=3, H=0.95; GFAP:
227 p.adjust=0.06, df=3, H=10.38; S100B: p.adjust=0.11, df=3, H=9.05). E, Averaged
228 trace of ATP-induced Ca^{2+} response assayed using FLIPR. Drugs or DMSO were
229 applied at 1 minute of the assay. The line represents the average fluorescence change
230 (ΔF/F0) in at least three independent experiments each with at least three replicate
231 wells. The shaded area represents the SEM across at least three independent
232 experiments. F, Quantitative comparison of the peak amplitude of ATP-induced Ca^{2+}
233 response among conditions (two-way ANOVA, p<2.2×10^{-16}, df=2, effect size=2.54)
234 and samples (p=2.87×10^{-14}, df=3, effect size=2.17). Error bars represent the SEM
235 across at least independent experiments. G: Quantitative comparison of the rise time
236 of ATP-induced Ca^{2+} response among conditions (two-way ANOVA, p=2.19×10^{-13},
237 df=2, effect size =1.958) and samples (p=0.064, df=3, effect size=0.76). Intergroup
238 comparison was performed using Post-hoc Tukey test. Error bars represent the SEM
239 across at least three independent experiments. (****: p<0.0001, ***: p<0.001, **:
240 p<0.01, *: p<0.05, ns: not significant).

241

Single cell RNA sequencing confirms the authenticity of PSC-derived astrocytes.

244 To further characterize the PSC-derived astrocytes, we performed full-length
245 scRNAseq on early and late BFP^{+} and BFP^{-} astrocytes using the iCELL8 platform and
246 SMART-seq technology, with non-patterned astrocytes derived from the LMX1A-
247 Cre/AAVS1-BFP tracer line as a control. A sample of iPSC-derived neurons was
248 included to facilitate downstream cell type identification. We profiled FACS purified
249 astrocytes expressing CD49f as well as unsorted cultures for comparison of all three
250 astrocyte populations (Barbar et al. 2020); Figure S4C-H). After stringent filtering
251 (Figure S5A-C, see Methods and Materials on filtering), we obtained a total of 17478
252 protein-coding genes in 1786 qualifying cells, with an average of 6326 protein-coding
253 genes detected per cell.

254 Unsupervised Louvain clustering identified 12 cell clusters (Figure 3A). Cells clustered
255 mainly based on sample type (astrocytes and neurons; Figure S6A) and the estimated
256 cell cycle phase (Figure S6B), while sorted CD49f^{+} and unsorted astrocytes were

257   clustered together (Figure S6A). *TagBFP* was detected at a higher level in BFP$^+$
258   astrocyte samples than in BFP$^-$ and NP samples, while *TagBFP* expression was
259   negligible in the neuronal samples derived from an iPSC line without BFP transgene
260   (Figure S6C-D). Using a set of known astrocyte and neuronal signature genes (Figure
261   3B), we identified cells in clusters 0, 1, and 5-11 as astrocytes (Figure 3B), which were
262   enriched in expression of *SOX9*, *NFIA*, *NFIB*, *HES1*, *S100A13*, *S100A16*, *EGFR*,
263   *CD44* and *GJA1* (Figure 3B). These transcripts were also detected at high levels in
264   clusters 2 and 4, which were mostly estimated to be in cell cycle phases G2, M and S
265   (Figure S6B). In addition, clusters 2 and 4 showed high levels of proliferation-related
266   transcripts, such as *TOP2A, MKI67, CDK1* and *AURKA* (Figure 3B), and are thus
267   defined as astrocyte precursors. In contrast, Cluster 3 contains mostly cells from the
268   neuronal sample (Figure 3A and Figure S6A) and indeed expressed high levels of
269   genes closely related to neuronal structure and function (such as *STMN2*, *SYT1*, *DCX*,
270   *MAPT*, and *SNAP25*; Figure 3B). We did not detect transcripts indicative of endoderm
271   (*GATA4*), mesoderm (*TBXT* and *TBX6*), and oligodendrocyte progenitors (*SOX10* and
272   *PDGFRA*) in any of these clusters (Figure S6C).

273   To determine the authenticity of these PSC-derived astrocytes, we mapped our data
274   to published scRNAseq datasets obtained from five foetal, an adult human brain and
275   a PSC-derived astrocyte study using Seurat integration (La Manno et al. 2016; Sloan
276   et al. 2017; Zhong et al. 2018; Polioudakis et al. 2019; Agarwal et al. 2020; Fan et al.
277   2020; Bhaduri et al. 2021; Eze et al. 2021). We found that cells in clusters annotated
278   as astrocytes (clusters 0, 1, and 5-11) were indeed predominantly mapped to the
279   reference astrocyte or astrocyte precursor populations with high confidence
280   (prediction score over 0.5; Figure 3C and Figure S7). In contrast, neuronal cluster 2
281   was mapped to neurons of the foetal reference datasets, while the astrocyte precursor
282   clusters (2 and 4) were mapped to progenitor populations in the foetal reference
283   datasets (Figure 3C and Figure S7). These findings demonstrate that our iPSC-
284   derived astrocytes closely resemble those in the human brains.

285
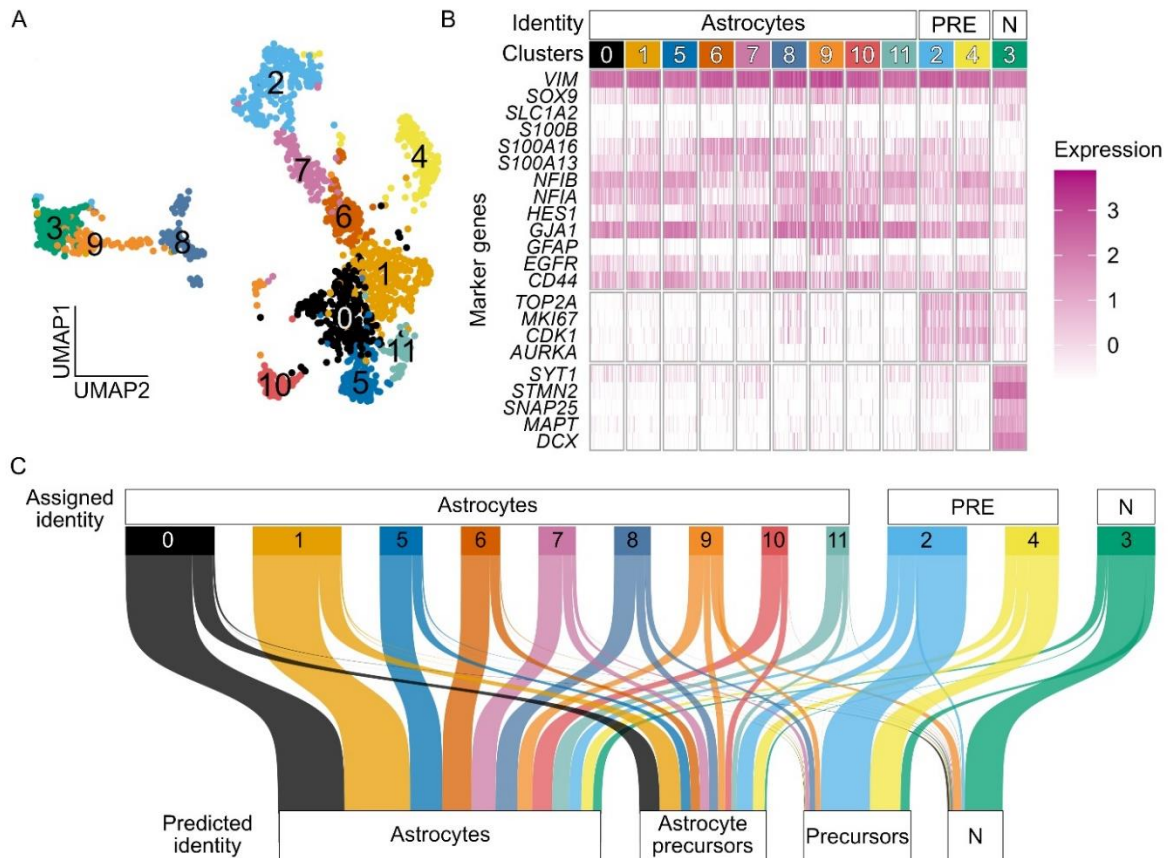
Figure 3. Single cell RNA sequencing confirms the authenticity of PSC-derived astrocytes.

A, Uniform manifold approximation and projection plot of unbiased clustering, coloured by clusters. B, Heatmap of the normalised expression of selected markers in different clusters. The assigned identity to each cluster is shown at the top of the plot. C, Sankey plot summarising the result of reference mapping of cells in different clusters to eight published reference human brain scRNAseq datasets. The thickness of the thread is proportional to the number of cells mapped to the same identity in the reference datasets (predicted identity). Detailed results of referencing mapping to each reference datasets are shown in Figure S7A-H and prediction score shown in Figure S7I. (PRE: precursors; N: neurons)

297

Distinct transcriptome fingerprints of LMX1A⁺ midbrain floor plate-derived astrocytes

Significant advance has been made recently in understanding the molecular profiles of midbrain dopamine neurons. However, our knowledge about midbrain astrocytes in this regard remains limited and does not inform anatomic or lineage origin of the cells.

13

303   In this regard, the BFP⁺ astrocytes provide a unique resource to determine the
304   transcriptomic characteristics of human midbrain floor plate derived astrocytes. By
305   performing pairwise differential gene expression (details described in Methods and
306   Materials), we identified 1153 genes differentially expressed (DEGs; adjusted p values
307   less than 0.05 and log2 fold change over 0.25) in BFP⁺ astrocytes when compared to
308   either BFP⁻ or non-patterned astrocyte populations (Supplementary Data 1). Of these,
309   159 were unique to BFP⁺ astrocytes (BFP⁺ enriched, Figure 4A), which include genes
310   associated with midbrain dopamine neuron development such as *SULF1*, *LMO3*,
311   *NELL2*, and *RCAN2* (Figure 4B) (Strelau et al. 2000; La Manno et al. 2016; Bifsha et
312   al. 2017; Ahmed et al. 2021). Interestingly, *LMX1A* and *FOXA2*, which was used to
313   evaluate PSC-derived midbrain astrocytes in previous studies, were not detected in
314   BFP⁺ astrocytes (Figure S6D). We also identified 530 DEGs enriched in BFP⁻
315   astrocytes only (BFP⁻ enriched, Figure 4A and Supplementary Data 1), which include
316   those known to be expressed in the ventrolateral - dorsal domain of the midbrain and
317   hindbrain, such as *IRX3*, *IRX5*, *PAX3*, and *PAX7* (Figure 4B)(Houweling et al. 2001;
318   Matsunaga et al. 2001). This transcription profile supports the notion that the BFP⁻
319   astrocytes are descendants of the initial minor populations of lateral midbrain
320   progenitors. Moreover, 72 DEGs were shared by BFP⁺ and BFP⁻ astrocytes against
321   the non-patterned astrocytes (midbrain enriched, Figure 4A and Supplementary Data
322   1). This set of genes includes *NR2F1*, *NR2F2, ZEB2, KCNJ6* and *SRPX* (Figure 4B)*,
323   which have been reported to be signatures of mouse midbrain astrocytes (Endo et al.
324   2022). Together, our findings provide a new entry to transcriptomic characteristics of
325   midbrain astrocytes and specifically a gene expression map of midbrain floor plate
326   derived human astrocyte lineage.

327   Gene ontology (GO) enrichment analysis was performed on the 1153 DEGs enriched
328   in BFP⁺ astrocytes (Supplementary Data 2). The significantly enriched GO terms were
329   mainly related to various aspects of metabolism, stress response, biosynthesis,
330   lysosomal activity, and cellular respiration (Figure 4C and Supplementary Data 3).
331   These biological processes have previously been shown to be disrupted by several
332   mutations causing familial Parkinson's disease (di Domenico et al. 2019; Barbuti et al.
333   2020; Sonninen et al. 2020). In contrast, the GO terms associated with 530 BFP⁻
334   astrocyte enriched DEGs were mostly related to formation of extracellular matrix and

335 tissue development (Figure 4D and Supplementary Data 2-3). The differential

336 enrichment of GO terms implicates functional difference between the BFP+ and BFP-

337 astrocytes and supports the need for generating regional specific astrocytes for

338 disease modelling.

339

340 **Figure 4. Distinct transcriptome fingerprints of LMX1A+ midbrain floor plate-**

341 **derived astrocytes**

342 A, Heatmap of the normalised expression of population-specific genes in different
343 populations of astrocytes. B, Violin plots of the normalised expression of selected
344 candidate markers for BFP+, BFP-, and non-patterned (NP) astrocytes. C-D,
345 Representative GO terms significantly enriched in BFP+ (C) and BFP- (D) enriched
346 genes. Semantically similar representative terms were shown with the same colour.

347

Discussion

349 Despite the general belief that recapitulating astrocyte lineage heterogeneity is
350 necessary for stem cell-based disease modelling and cell transplantation, the extent
351 of astrocyte heterogeneity in different brain regions and subregions remains largely
352 elusive. By harnessing an LMX1A based lineage tracing human iPSC line and cutting
353 edge scRNAseq technology, we show that astrocytes derived from the LMX1A+
354 midbrain floor plate progenitors, the same cells giving rise to midbrain dopaminergic
355 neurons, possess distinct transcriptional landmarks from those derived from non-
356 midbrain patterned neural progenitors as well as other midbrain patterned progenitors.
357 Moreover, we discovered unexpected negative selection against derivatives of
358 LMX1A+ progenitors during astrocyte induction and progenitor expansion. Our study
359 highlights the need for careful characterisation of PSC-derived astrocytes and
360 provides a transcriptomic fingerprint for midbrain floor plate derived astrocytes.

361 Using a popular astrocyte in vitro differentiation paradigm, we found that astrocytes
362 descended from the LMX1A+ midbrain progenitors could only be obtained from purified
363 progenitors. In contrast, astrocytes derived from bulk midbrain patterned progenitors
364 exhibits transcriptomic profiles of the lateral-dorsal midbrain, despite LMX1A+ midbrain
365 progenitors being the predominant starting population. Our findings demonstrate that
366 the lineage composition of the parent progenitors may not be faithfully preserved
367 during astrocyte induction and progenitor expansion. FGF is the most popular
368 inductive molecule used for astrocyte differentiation from stem cells (Chandrasekaran
369 et al. 2016). It is evident however that FGF expanded neural progenitors, originated
370 either from the brain or neutralized PSCs, exhibit restricted regional competence and

16

371 positional gene expression. For example, bulk expanded human ventral midbrain
372 neural progenitors (Jain et al. 2003), and fetal forebrain or spinal cord derived neural
373 stem (NS) cells only give rise to GABAergic neurons (Sun et al. 2008); likewise, lt-
374 NES cells display anterior hindbrain-like positional profile (Falk et al. 2012), while their
375 antecedents, PSC-derived neural rosettes and early passage derivatives, express
376 anterior forebrain markers (Koch et al. 2009). It is not clear whether this is due to
377 deregulation of the original patterning at the level of gene expression or the loss of
378 associated cell population (Gabay et al. 2003). In this study, since BFP⁺ astrocytes
379 can be generated under the same culture condition with purified LMX1A⁺ progenitors,
380 we reasoned that the loss of their derivatives in unsorted cultures was possibly due to
381 differential growth capacity.

382 Our study highlights the need for careful assessment of astrocyte positional identity.
383 A common practice in this regard is to confirm the regional characteristics of the
384 founder progenitors following fate directed neural induction, with the assumption that
385 the dominant positional features will be maintained by the astrocyte progeny (Krencik
386 et al. 2011). This strategy is at least partly dictated by our limited knowledge in gene
387 expression signatures of regional- and/or lineage-specific astrocytes. Hence, an end
388 point evaluation of PSC-derived astrocytes often relies on region-specific markers
389 defined in the developmental brain during the neurogenic period. For example, LMX1A
390 and FOXA2 expression were used as criteria for midbrain astrocytes in previous
391 studies (Barbuti et al. 2020; Crompton et al. 2023). However, scRNAseq of human
392 fetal ventral midbrain and adult substantia nigra revealed negligible expression of
393 these transcripts in astrocytes (La Manno et al. 2016; Agarwal et al. 2020; Kamath et
394 al. 2022). Consistent with these findings, we did not detect LMX1A or FOXA2 in neither
395 the BFP⁺ nor BFP⁻ astrocytes. Our analysis however identified new positive and
396 negative markers that could be applied to confirm the regional identity of ventral
397 midbrain astrocytes.

398 In addition to distinct transcriptomic profile, BFP⁺ and BFP⁻ astrocytes may also be
399 functionally different. Astrocytes generated from progenitors broadly patterned to the
400 dorsal forebrain, ventral forebrain and spinal cord have been shown to exhibiting
401 different GO enrichment profile as well as different physiological and functional
402 properties (Bradley et al. 2019). In comparison to the BFP⁻ and non-patterned

403 astrocytes, the current study revealed that GO terms enriched in BFP$^+$ astrocytes,
404 which originated from the same progenitor giving rise to midbrain dopaminergic
405 neurons, were closely related to various biological processes disrupted in astrocytes
406 carrying familial Parkinson's disease mutations (di Domenico et al. 2019; Barbuti et al.
407 2020; Sonninen et al. 2020). Such a distinct enrichment profile implicates BFP$^+$
408 astrocytes being functionally adapted to supporting midbrain dopaminergic neurons
409 compared to BFP$^-$ and non-patterned astrocytes.

410 In conclusion, this study lends further support on regional diversity of astrocytes and
411 identified a set of midbrain enriched genes. Crucially, the transcriptomic fingerprint of
412 human midbrain floor plate-derived astrocytes described here offers a much-needed
413 resource for assessing the authenticity of stem cell derived astrocytes in studies
414 associated with Parkinson's disease.

Methods and Materials

Stem cell culture and astrocyte differentiation

417 KOLF2 human iPSCs were maintained in E8 flex media (ThermoFisher) and manually
418 dissociated using Gentle Cell Dissociation Reagent (STEMCELL Technologies) as
419 previously described (Cardo et al. 2023). Astrocytes were differentiated using a three-
420 stage stepwise strategy consisting neural induction and regional patterning, astrogenic
421 switch and progenitor expansion, and astrocyte terminal differentiation. Midbrain floor
422 plate progenitors were generated as previously described (Cardo et al. 2023). At day
423 19, cells were replated as single cells onto poly-D-lysine-laminin-coated plates at
424 1×10^6 cells/cm^2 for astrogenic switch and progenitor expansion in N2B27 media
425 supplemented with 10 ng/mL FGF2 (Peprotech) and 10 ng/mL Human EGF
426 (Peprotech) and replated every 6-8 days. For astrocyte terminal differentiation,
427 expanded neural progenitors were re-plated at a density of 3×10^4 cells/cm^2 in
428 expansion media and 24 hours later switched to N2B27 supplemented with 10 ng/mL
429 human recombinant CNTF (Peprotech) and 10 ng/mL human recombinant BMP4
430 (Peprotech) for 7 days followed by media containing CNTF alone for another 13 days.
431 10 µM Y-27632 was used for 24 hours before and after each replating. The protocol

432   for generating non-patterned astrocytes was the same as for floor plate-derived

433   astrocyte except the neural progenitors were derived with duo-SMAD inhibitors only

434   without ventral patterning reagents.

Flow cytometry analysis and cell isolation

436   Cells were dissociated in Accutase as described above and washed twice with DPBS

437   by centrifugation for 5 minutes at 200 rcf. For evaluating BFP expression, dissociated

438   cells were resuspended in 0.5 mM EDTA in DPBS (Sigma-Aldrich) and analysed on a

439   BD LSRFortessa cell analyser (BD Biosciences). For purifying BFP$^+$ cells, dissociated

440   cells were resuspended in the same cell culture media. Background autofluorescence

441   was compensated for using KOLF2 parental cell line at a similar stage of differentiation

442   to define BFP$^-$ gating. For purifying CD49f$^+$ astrocytes, dissociated cells were stained

443   with Alexa Fluor 647-conjugated rat anti-CD49f antibody (5% v/v in a 100 μL reaction;

444   BD Biosciences) for 25 minutes at 37°C on an orbital shaker at 200 rcf and

445   resuspended in DPBS containing 0.5% bovine serum albumin and 50 units/mL DNase

446   I (Sigma Aldrich). Background autofluorescence was compensated for using KOLF2

447   parental cell line at a similar stage of differentiation to define BFP$^-$ gating and unstained

448   astrocytes to define CD49f$^-$ gating. Cell sorting was performed on a BD FACSAria III

449   (BD Biosciences) using an 80 μm nozzle. Sorted cells were collected in the same

450   media as for resuspension. Flow cytometry data were analysed in FlowJo v10.8.1 (BD

451   Biosciences) as shown in Figure S1B-E. Briefly, non-debris events were selected

452   using the eclipse gates on dot graphs of SSC-A versus FSC-A. Singlet events were

453   sequentially gated using polygonal gates on dot graphs of FSC-H versus FSC-A and

454   SSC-H versus SSC-A by selecting the events in the diagonal region. The positive and

455   negative gates in the fluorescence channel were set as bifurcate gates at a minimum

456   of 99.9% percentile (usually at 99.99% percentile) on the histogram of the

457   fluorescence intensity of the negative control sample of the same flow cytometry

458   experiment and applied to all samples of the same flow cytometry experiment.

Immunocytochemistry

460   Cultures were fixed with 3.7% PFA for 15-20 min at 4 ºC. For nuclear antigen

461   detection, an additional fixation with methanol gradient was performed, which include

462 5 mins each in 33% and 66% methanol at room temperature followed by 100%
463 methanol for 20 min at -20°C. Cultures were then returned to PBST via inverse
464 gradient and were then permeabilized with three 10-minute washes in 0.3% Triton-X-
465 100 in PBS (PBS-T) and then blocked in PBS-T containing 1% BSA and 3% donkey
466 serum. Cells were incubated with primary antibodies in blocking solution overnight at
467 4°C. Following three PBS-T washes, Alexa-Fluor secondary antibodies (Thermo
468 Fisher Scientific) were added at 1:1000 PBS-T for 1 hour at ambient temperature in
469 the dark. Three PBS-T washes were then performed that included once with DAPI
470 (Molecular Probes). Images were taken on a Leica DMI6000B inverted microscope.
471 Quantification was carried out in Cell Profiler (Stirling et al. 2021) or manually using
472 ImageJ (Schindelin et al. 2012) by examining at least four randomly selected fields
473 from three independent experiments. The antibodies used are provided in the
474 Supplementary Table 1. Representative images shown in main figures were cropped
475 by randomly selecting the region of interest in the DAPI-stained channel only, with the
476 original unedited images shown in Figure S1A, S2 and S3.

Single-cell RNA-sequencing

478 Cells were dissociated with Accutase with 10 units/mL of papain (Sigma-Aldrich) for
479 10 minutes at 37°C and resuspended in 0.5% bovine serum albumin (Sigma Aldrich)
480 with 50 units/mL DNase I in DPBS without calcium or magnesium (Gibco) and passed
481 through a cell strainer with 35 µm mesh. Cells were stained with 1 µM SYTO16
482 (Invitrogen) and 0.08% (v/v) propidium iodide (part of the Invitrogen ReadyProbes™
483 Cell Viability Imaging Kit, Blue/Red) for 20 minutes on ice and then dispensed into the
484 nano-well plates of the ICELL8® cx Single-Cell System (Takara). Wells containing
485 single viable cells were automatically selected using the ICELL8 cx CellSelect v2.5
486 Software (Takara) with the green and not red logic. Manual triage was performed to
487 recover additional candidate wells that contain viable single cells. The library for
488 sequencing was then prepared using the SMART-Seq ICELL8 application kit (Takara)
489 following the manufacturer's recommended protocol. Next-generation sequencing was
490 performed using the NovaSeq6000 and the Xp Workflow on a S4 flow cell for 200
491 cycles of pair-end sequencing.

Single-cell RNA-sequencing analysis

492

493 Using the Cogent NGS Analysis Pipeline software v 1.5.1 (Takara), FASTQ files

494 containing all indices for each chip were demultiplexed to FASTQ files containing one

495 index per file. Adaptor sequences were removed using cutadapt 3.2 with the following

496 settings: -m 15 --trim-n --max-n 0.7 -q 20. Trimmed FASTQ files were aligned to the

497 *Homo sapiens* GRCh38.106 primary assembly with the BFP reporter gene attached

498 to the end of the genome, using STAR 2.7.9a (Dobin et al. 2013) with the following

499 settings: --outSAMtype BAM Unsorted --quantMode TranscriptomeSAM --

500 outReadsUnmapped Fastx --outSAMstrandField intronMotif --chimSegmentMin 12 --

501 chimJunctionOverhangMin 8 --chimOutJunctionFormat 1 --alignSJDBoverhangMin 10

502 --alignMatesGapMax 100000 --alignIntronMax 100000 --alignSJstitchMismatchNmax

503 5 -1 5 5 --chimMultimapScoreRange 3 --chimScoreJunctionNonGTAG -4 --

504 chimMultimapNmax 20 --chimNonchimScoreDropMin 10 --peOverlapNbasesMin 12 -

505 -peOverlapMMp 0.1 –alignInsertionFlush Right --

506 alignSplicedMateMapLminOverLmate 0 --alignSplicedMateMapLmin 30. After

507 alignment, gene-level quantification was performed using featureCounts from subread

508 2.0.0 (Liao et al. 2013) with the following settings: -t exon --primary -R CORE -F GTF

509 -Q 0 -B -g gene_id. The count matrix of each index was combined in R 4.2.0 (Team

510 2023).

511 All downstream analysis was performed in R 4.3.0 using Seurat 4.3.0 (Stuart et al.

512 2019). Gene-level filtering was applied by including only protein-coding genes with at

513 least five total counts across all cells and being expressed in at least 1% of all cells.

514 Poor quality cells were then identified using the *is.outlier* function from the scater

515 1.28.0 (McCarthy et al. 2017). Poor quality cells were defined as having a high

516 percentage of mitochondrial gene count, or high or low the total number of genes

517 detected, or high total gene counts. The thresholds of each metrics for each sample

518 were determined as twice the median absolute deviation of the sample. Raw gene

519 counts were log normalised with a scale.factor setting of 1×10^5. Data from the two

520 batches of experiments were integrated using the *FindIntegrationAnchors* and

521 *IntegrateData* based on the common top 2000 highly variable genes and the first 30

522 dimensions of principal components. The percentage of mitochondrial gene count and

21

523   total gene count were regressed out using the *ScaleData* function. Principal
524   component analysis (PCA) was performed on the top 2000 high variable genes and
525   the number of principal components used for uniform manifold approximation and
526   projection (UMAP) was determined using the JackStraw method (Chung and Storey
527   2015). UMAP and unbiased Louvain clustering was performed on the first 33 principal
528   components. Pairwise differential gene expression analysis was performed using the
529   MAST method (Finak et al. 2015) with "Chip" as the latent variable. Gene ontology
530   enrichment analysis was performed using *enrichGO* function in the clusterProfiler
531   4.10.0 package (Stirling et al. 2021) with all genes in the filtered dataset as the
532   background. GO term database was downloaded using the org.Hs.eg.db 3.18.0
533   package (Carlson 2019). Revigo v1.8.1 was used to group the representative GO
534   terms based on semantic similarity using a size setting of 0.5, Homo sapiens
535   database, and SimRel method for semantic similarity (Schlicker et al. 2006; Supek et
536   al. 2011).

537   Published datasets were downloaded from NCBI's Gene Expression Omnibus
538   (Clough and Barrett 2016) and processed in R 4.2.0. Gene level filtering was
539   performed by retaining only protein-coding genes with more than five total counts
540   across all cells. Gene counts were normalised using the NormalizeData function
541   (scale.factor settings are listed in Supplementary Table 2). PCA was performed based
542   on the top 2000 highly variable genes to obtain the first 50 PCs. Visual inspection of
543   the elbow plot was used to determine the number of PCs for downstream analysis.
544   Batch effect between subjects was evaluated on the two-dimensional PC2~PC1 plot.
545   Where inter-subject batch effect was observed, Harmony integration was performed
546   based on the PCs selected in the previous step (Supplementary Table 2). UMAP was
547   performed based on either the PCA or Harmony reduction (using the top 30
548   dimensions), and Louvain clustering was performed (settings shown in Supplementary
549   Table 2). Cluster identities were verified against the reported annotation where
550   possible. For datasets without detailed annotation published or astrocyte lineage
551   reported (Supplementary Table 2), reannotation was performed based on the
552   expression of known markers (Figure S7A-E). Reference mapping was performed
553   using *FindTransferAnchors* and *TransferData* function in Seurat based on the first 30
554   dimensions of either the PCA or Harmony loadings of the reference dataset.

Statistical analyses

All data were collected from at least three independent experiments and presented as mean ± standard error of means unless otherwise specified. Data were tested for normality with the Shapiro-Wilk test and for equal variance with Levene test before performing statistical analyses by two-way ANOVA with post-hoc Tukey test for multiple comparisons where relevant. Kruskal-Wallis test with post-hoc Dunn's test for pairwise comparison was used where parametric test was not suitable. Effect size was calculated as Cohen's f for ANOVA or eta squared based on the H-statistic for Kruskal-Wallis test. All statistical tests were performed in R4.3.0.

23

564 **Supplementary Information**
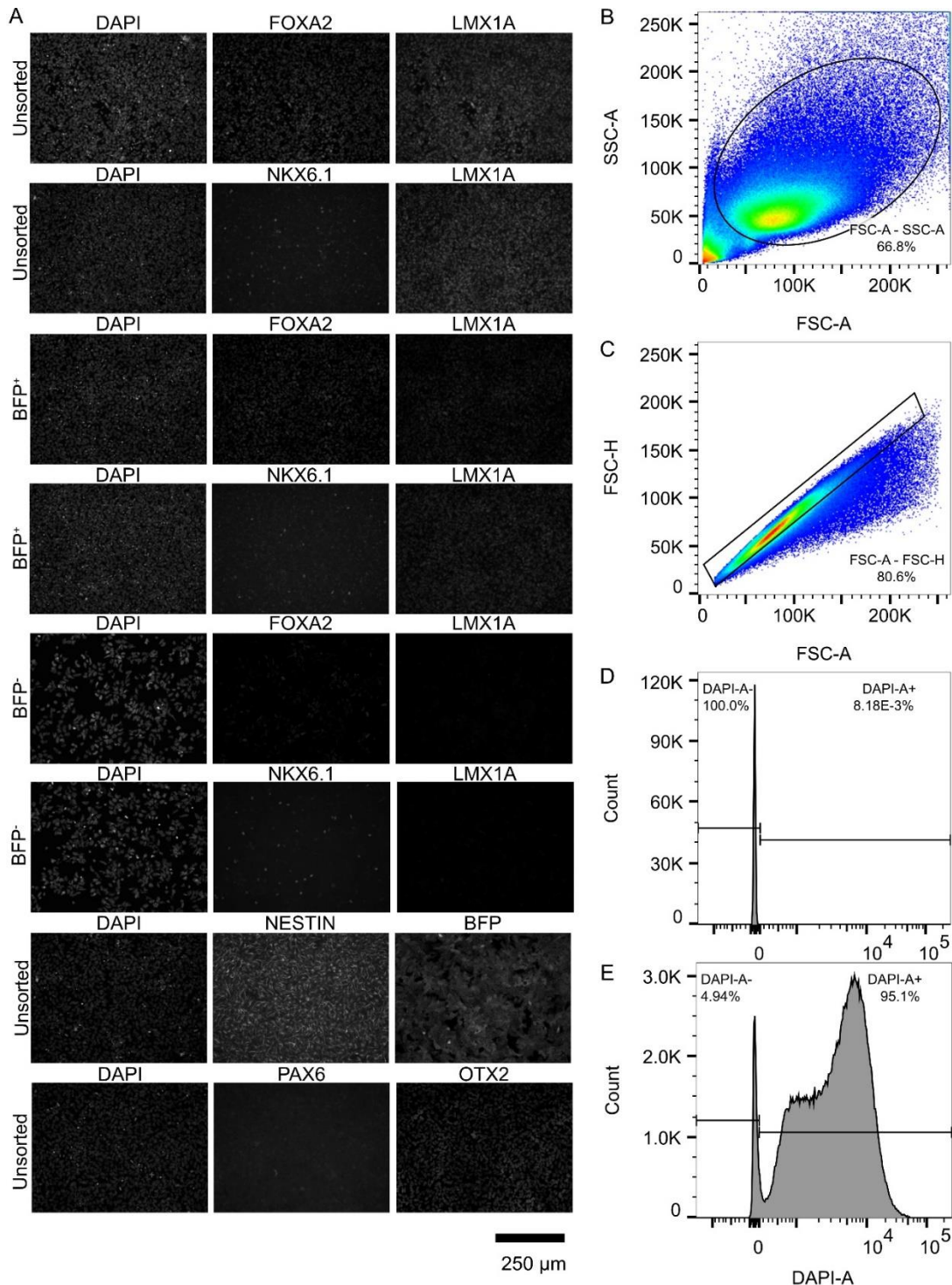
Figure S1. Original images of immunocytochemistry of d19 progenitors and gating strategy of BFP flow cytometry analysis.

A, the original images shown in Figure 1B-C. The gating strategy used for BFP flow cytometry analysis are shown in B-E and described in Methods and Materials. B,

570 scatter plot of SSC-A versus FSC-A. C, scatter plot of FSC-H versus FSC-A. D-E,

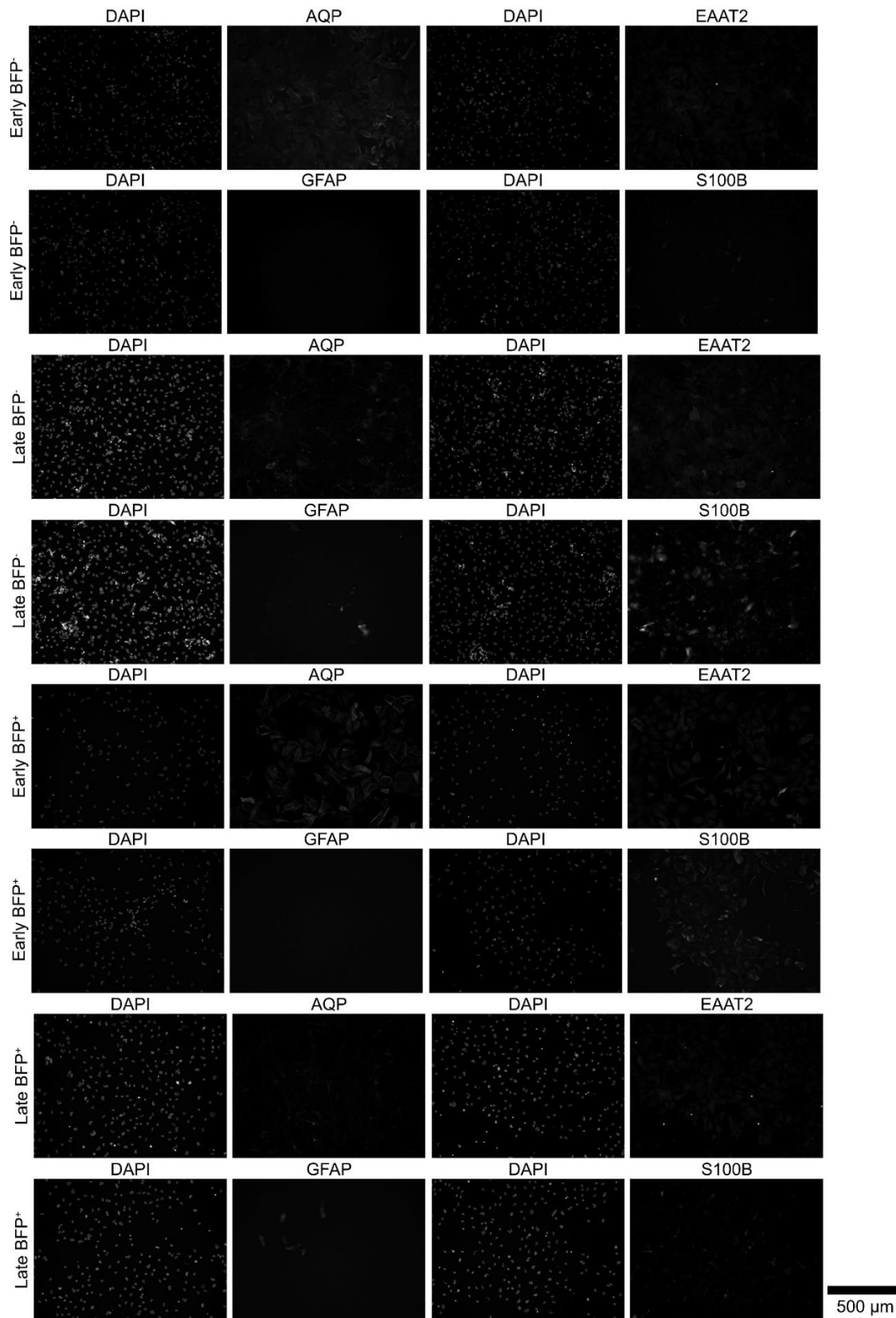571 histogram of BFP fluorescence in the negative control (D) and BFP-expressing (E)

572 samples.

573
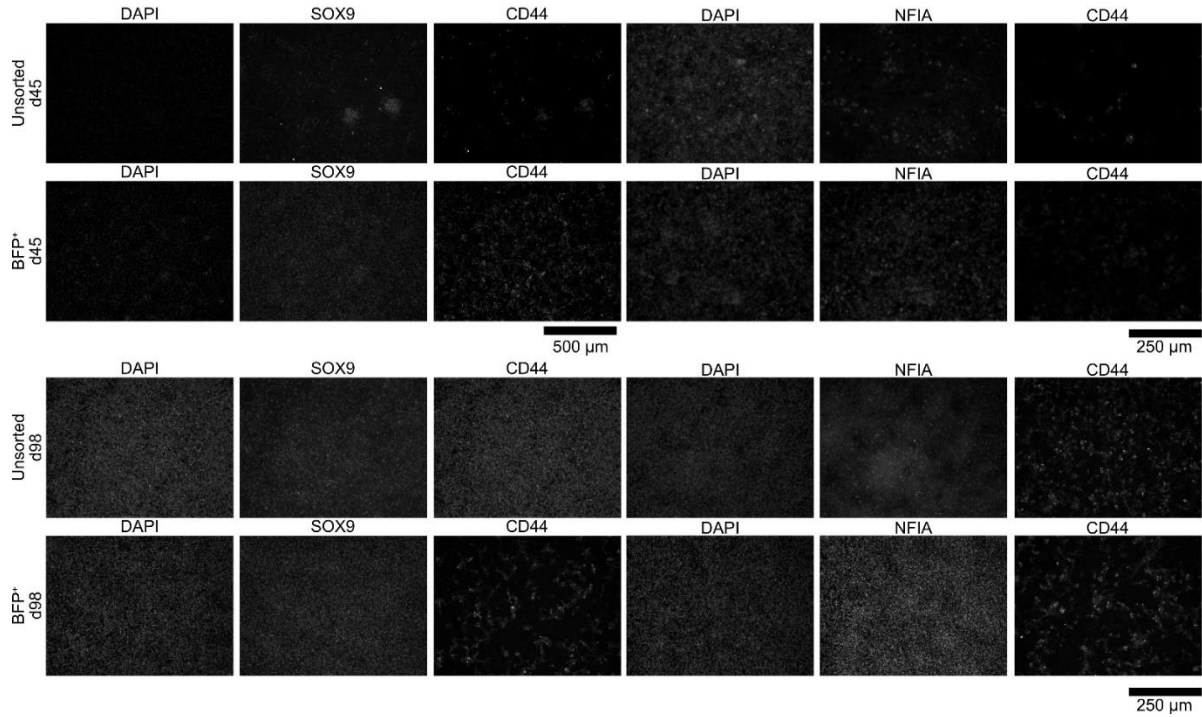
574

575 **Figure S2. Original images of immunocytochemistry of astrogenic markers**

576 **shown in Figure 2A.**

577

**Figure S3. Original images of immunocytochemistry of astrocyte markers shown in Figure 2B.**

580
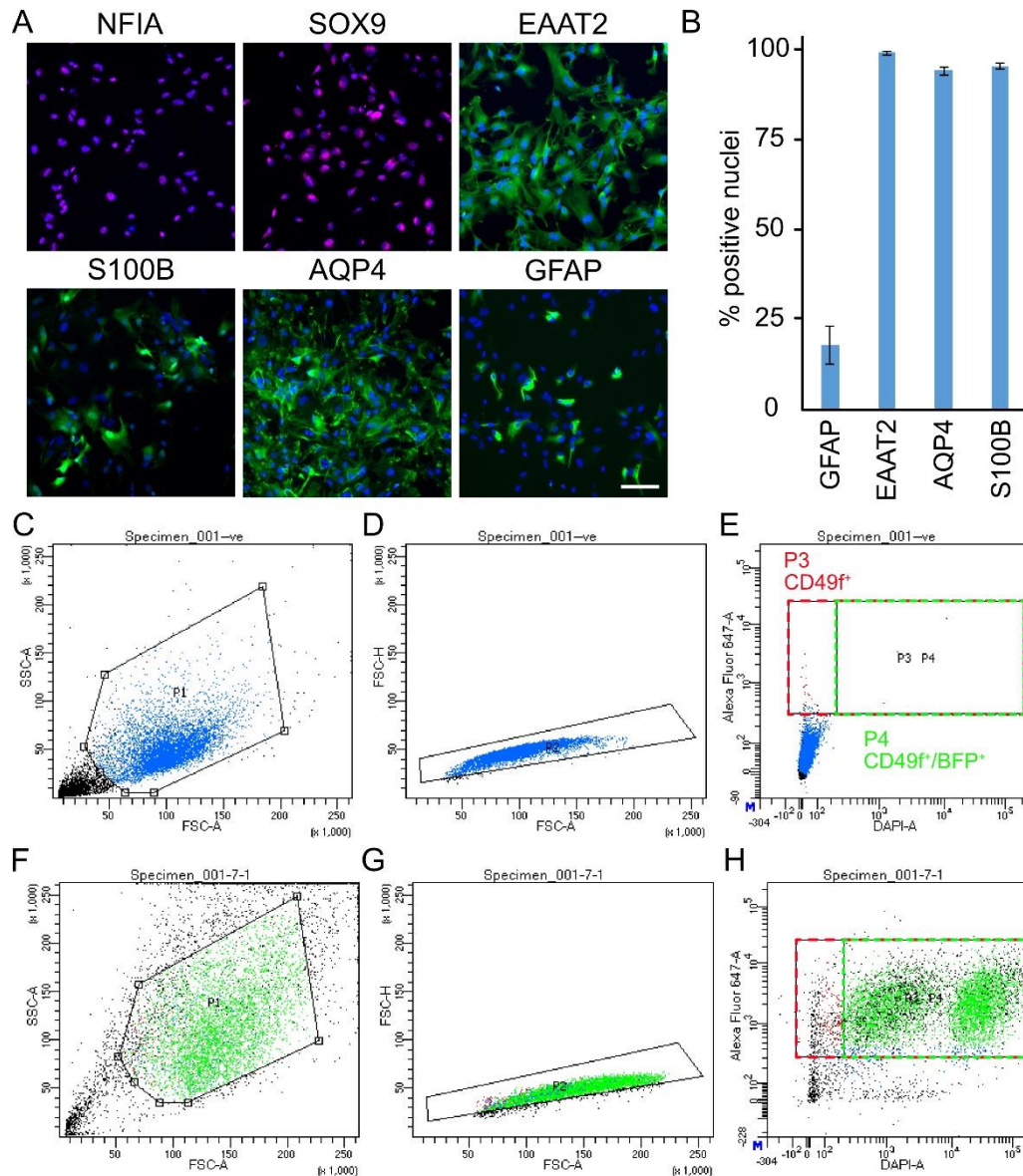
**Figure S4. Non-patterned astrocytes and fluorescence-activated cell sorting of astrocytes for single cell RNA sequencing.**

A, Representative view of immunocytochemistry of astrocyte marker expression in non-patterned astrocytes (scale bar represents 100 μm). B, Quantification of astrocyte marker expression in astrocytes. Error bars represent SEM of three independent experiments. C-E, negative control samples; F-H, one sample of BFP+ astrocytes). C and F shows the dot plot of SSC-A versus FSC-A. D and G shows the dot plot of FSC-H versus FSC-A. E and H shows the dot plot of Alexa Fluor-647-A (labelling CD49f) versus DAPI-A (labelling BFP). P3 was used to isolate CD49f$^+$ population (including both BFP$^+$ and BFP$^-$), while P4 was used to isolate CD49f$^+$/BFP$^+$ population.

591
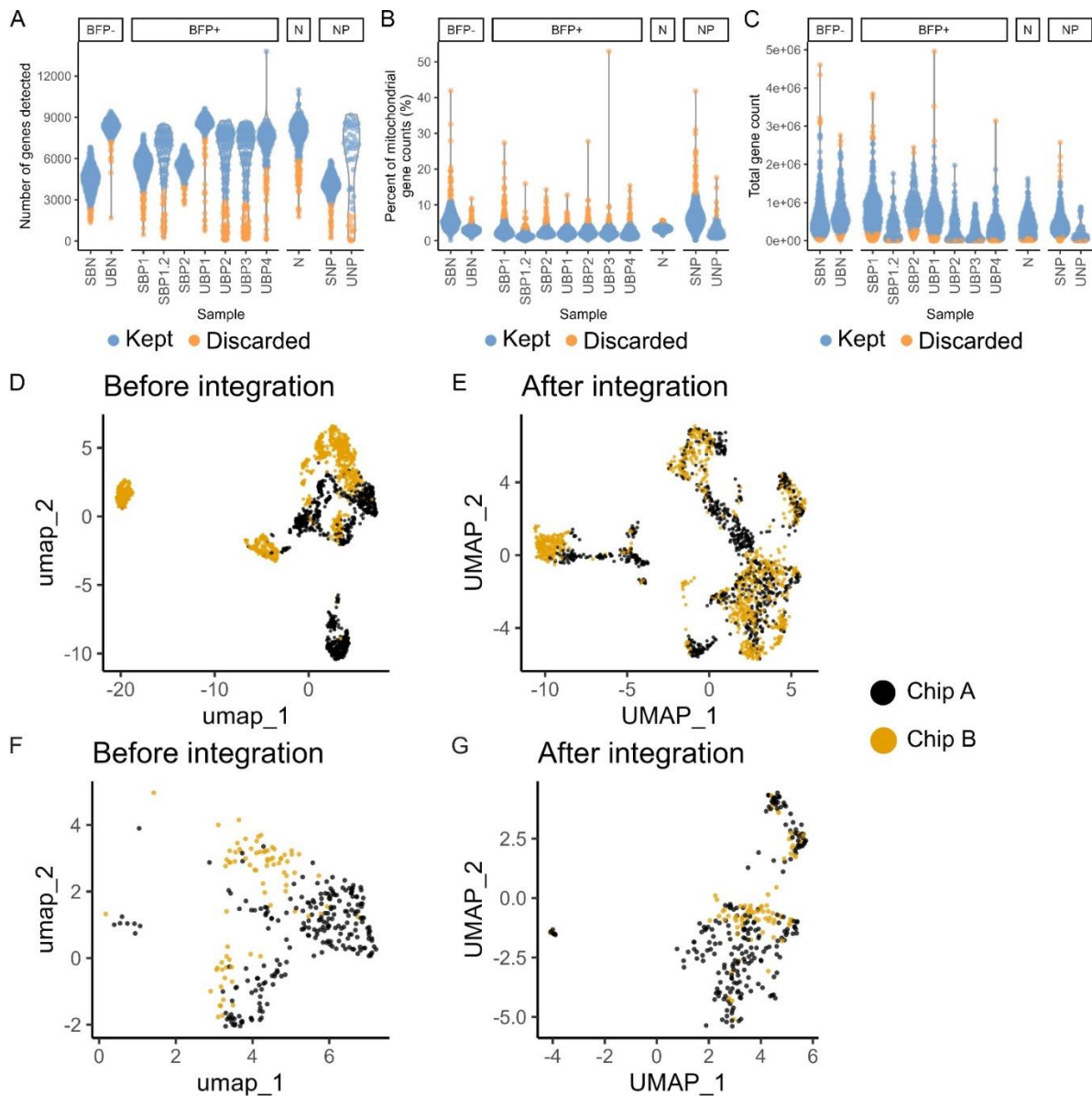
Figure S5. Processing of single cell RNA sequencing data.

A-C, Violin plots showing the results of adaptive cell-level filtering based on the number of genes detected per cell (A), percentage of mitochondrial gene counts per cell (B), and total gene count per cell (C). D-E, UMAP plot of all filtered cells before (D) and after integration (E) coloured by chip. F-G, UMAP plot of the subset of sorted BFP+ astrocytes on Chip A and B derived from the same astrocyte differentiation before (F) and after integration (G) coloured by chip. (N: neuron; SBN: sorted BFP-; SBP: sorted BFP+; SNP: sorted non-patterned; UBN: unsorted BFP-; UBP: unsorted BFP+; UNP: unsorted non-patterned; numbers represent independent samples)

28
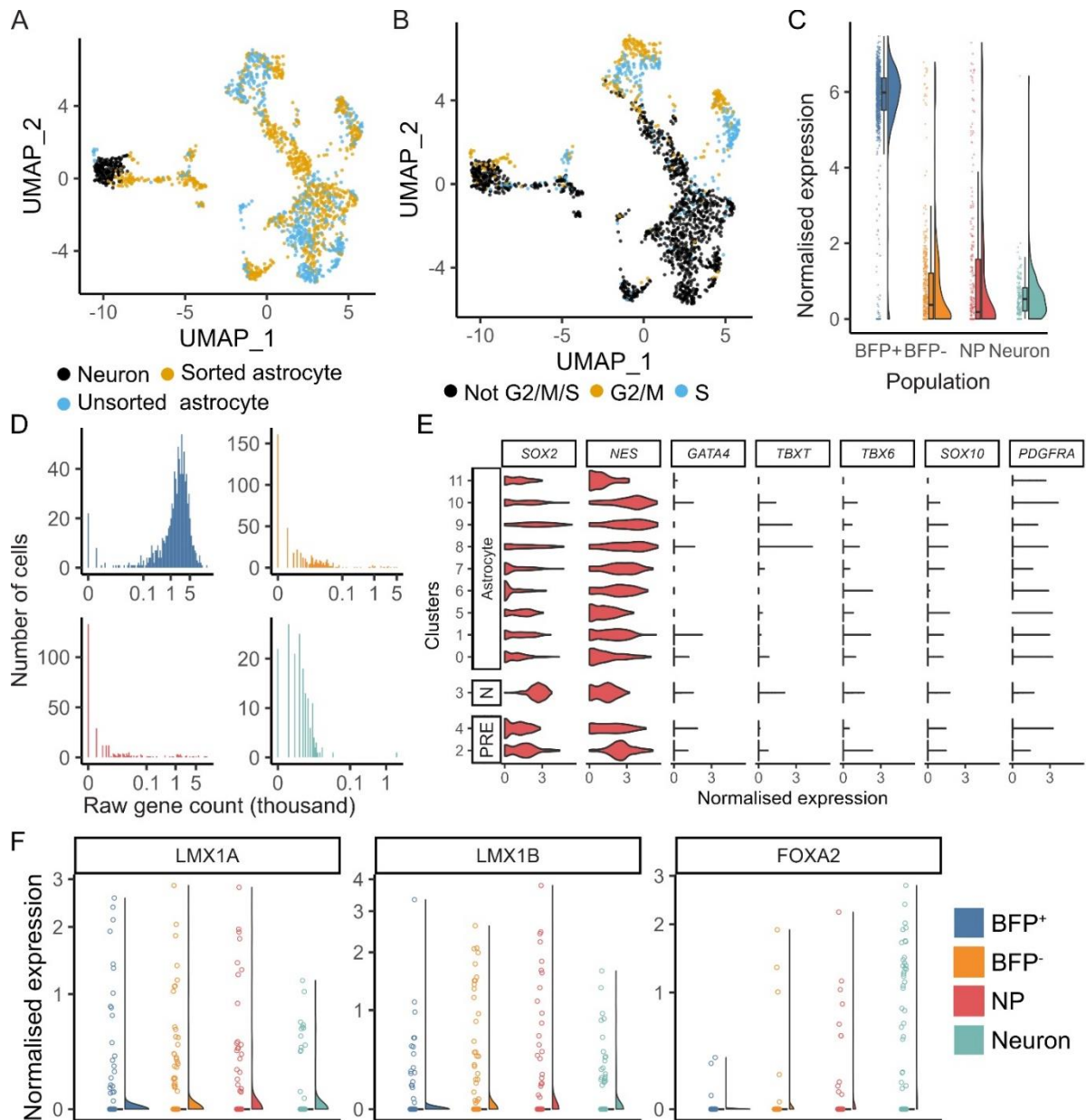
Figure S6. Marker expression in scRNAseq.

A-B, UMAP plot of all filtered cells after integration coloured by sample type and sorting status (A) and estimated cell cycle phase (B). C, Raincloud plots of normalised expression of *BFP*. D, Histogram of the distribution of raw gene count of *BFP*. E, Violin plots of the normalised expression of endoderm, mesoderm, neuroectoderm, and oligodendrocyte progenitor markers. F, Violin plots of the normalised expression of classic ventral midbrain markers.
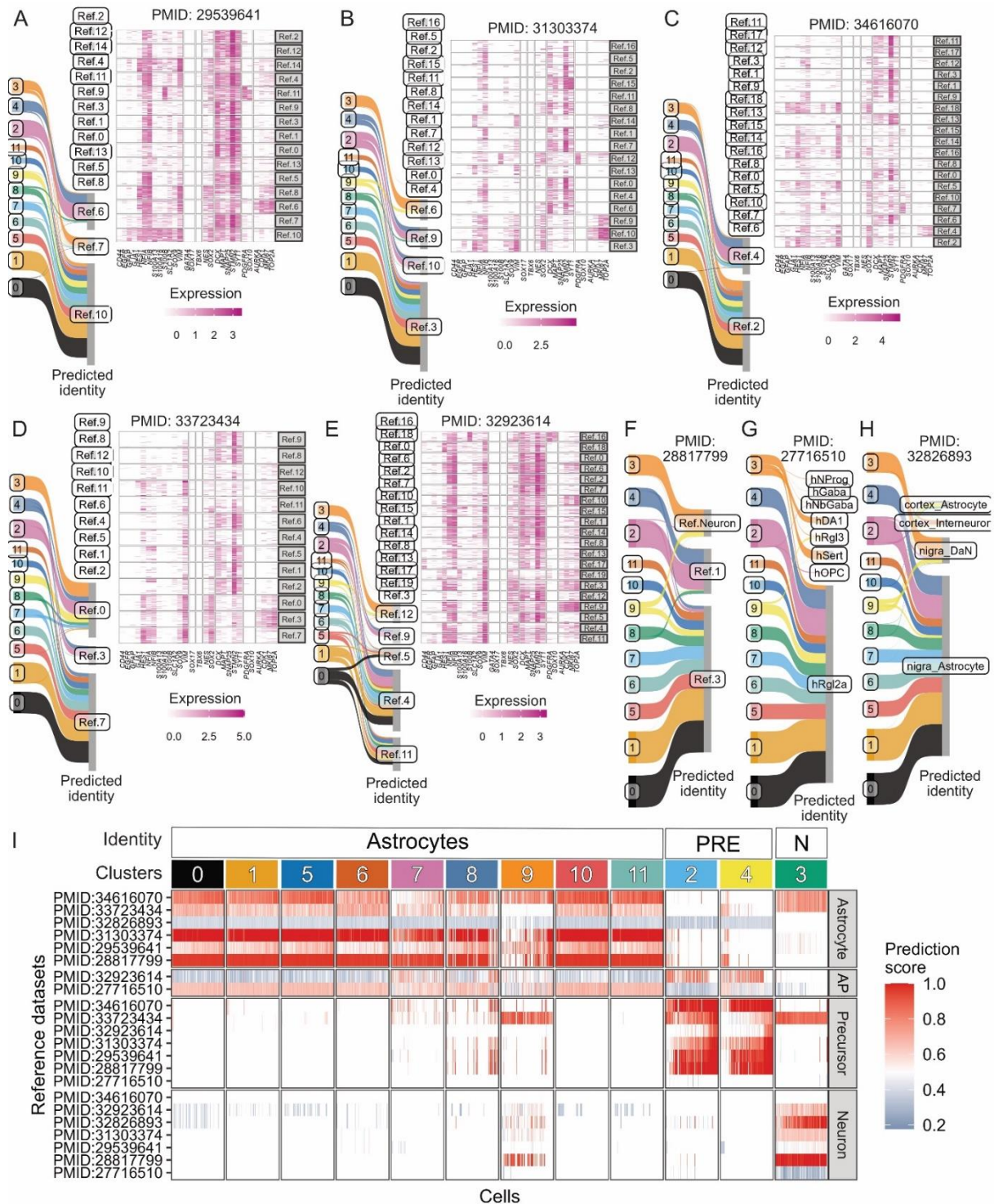
Figure S7. Reference mapping to human brain single cell RNA sequencing datasets.

A-H, Sankey plot summarising the result of reference mapping of cells in different clusters to eight published reference human brain scRNAseq datasets. The thickness of the thread is proportional to the number of cells mapped to the same identity in the

30

615   reference datasets (predicted identity). Cluster IDs in this study are shown on the left.

616   Heatmap in Panel A-E shows the expression of marker genes in different clusters in

617   the re-annotated reference datasets. I, Heatmap of prediction score from Seurat

618   integration.

619

620   **Supplementary Table 1. Antibodies used in this study.**

Target	Species	Producer	Catalogue number	Dilution for use
ALDH1L1	MS	Abcam	ab56777	1:500
AQP4	RB	Abcam	ab259318	1:200
BFP	rabbit	Evrogen	AB233	1:3000
CD44	mouse	Cell Signalling Technology	3750	1:400
EAAT2	goat	Abcam	ab235202	1:200
FOXA2	goat	R&D system	AF2400	1:1000
GFAP	rat	Life Tech	AB10533	1:1000
LMX1A	rabbit	Millipore	HPA030088	1:2000
NESTIN	mouse	BD	BD611659	1:300
NFIA	rabbit	Abcam	AB228897	1:1000
NKX6.1	mouse	DSHB	F55A10	1:100
OTX2	goat	R&D system	AF1979	1:200
PAX6	mouse	DSHB	PAX6	1:1000
S100B	rabbit	Abcam	ab52642	1:1000
SOX9	goat	R&D system	AF3075	1:300
Anti-Goat IgG, Alexa Fluor 488	donkey	Invitrogen	A32814	1:1000
Anti-Goat IgG, Alexa Fluor 555	donkey	Invitrogen	A32816	1:1000
Anti-Goat IgG, Alexa Fluor 647	donkey	Invitrogen	A32849	1:1000

31

Anti-Mouse IgG, Alexa Fluor 488	donkey	Invitrogen	A21202	1:1000
Anti-Mouse IgG, Alexa Fluor 555	donkey	Invitrogen	A31570	1:1000
Anti-Mouse IgG, Alexa Fluor 647	donkey	Invitrogen	A31571	1:1000
Anti-Rabbit IgG, Alexa Fluor 488	donkey	Invitrogen	A21206	1:1000
Anti-Rabbit IgG, Alexa Fluor 555	donkey	Invitrogen	A31572	1:1000
Anti-Rabbit IgG, Alexa Fluor 647	donkey	Invitrogen	A31573	1:1000
Anti-Rat IgG, Alexa Fluor 488	donkey	Invitrogen	A21208	1:1000

621

622 **Supplementary Table 2. Settings used for processing published datasets.**

PMID	Tissue	Age	Scale factor	PCs used	Harmony integration	Clustering resolution	Re-annotation
3282 6893	SN and cortex	Adult	10000	50	No	0.5	No
3372 3434	multiple	CS22	10000	50	Yes	0.3	Yes
3461 6070	multiple	GW25	10000	50	Yes	0.4	Yes
3292 3614	Neocortex	GW9-28	100000	30	Yes	0.9	Yes
3130 3374	Neocortex	GW17-18	10000	30	Yes	0.5	Yes
2953 9641	Prefrontal cortex	GW8-26	100000	30	Yes	0.9	Yes
2881 7799	PSC-derived astrocytes	-	10000	30	No	0.3	No
2771 6510	Ventral midbrain	GW7-11	10000	30	No	0.5	No

623

624 Supplementary Data 1: Pairwise DEGs of three astrocyte and one neuron populations.

625 Supplementary Data 2: GO enrichment of population enriched DEGs.

626 Supplementary Data 3: Representative enriched GO terms of DEGs for BFP+ and
627 BFP- astrocytes.

628 Supplementary Data 4: DEGs comparing sorted and unsorted populations.

Acknowledgements

Author Contribution

636 Z.L. and M.L. conceived the study and designed the experiments. Z.L. performed all
637 cell experiments and analysed all data. L.C. and M.L. designed the lineage tracing
638 system, and L.C. generated the cell line with assistance from Z.L. M.R. performed
639 iCELL8 library preparation. J.M.S., F.W., and V.V. provided guidance and contributed
640 to general discussions on scRNAseq data analysis. C.W. contributed to the design of
641 scRNAseq and general discussions throughout the work. Z.L. and M.L wrote the
642 paper. All authors edited and approved the paper.

Funding

Data Availability

The sequencing data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Clough and Barrett 2016) and are accessible through a GEO Series accession number which will be made available upon final publication.

Declaration of interests

The authors declare no competing interests.

References

Agarwal, D. et al. 2020. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat Commun* 11(1), p. 4183. doi: 10.1038/s41467-020-17876-0

Ahmed, M., Owens, M. J. S., Toledo, E. M., Arenas, E., Bradley, M. and Ffrench-Constant, C. 2021. Combinatorial ECM Arrays Identify Cooperative Roles for Matricellular Proteins in Enhancing the Generation of TH+ Neurons From Human Pluripotent Cells. *Front Cell Dev Biol* 9, p. 755406. doi: 10.3389/fcell.2021.755406

Andersson, E. et al. 2006. Identification of intrinsic determinants of midbrain dopamine neurons. *Cell* 124(2), pp. 393-405. doi: 10.1016/j.cell.2005.10.037

Barbar, L. et al. 2020. CD49f Is a Novel Marker of Functional and Reactive Human iPSC-Derived Astrocytes. *Neuron* 107(3), pp. 436-453 e412. doi: 10.1016/j.neuron.2020.05.014

Barbuti, P. A. et al. 2020. IPSC-derived midbrain astrocytes from Parkinson's disease patients carrying pathogenic SNCA mutations exhibit alpha-synuclein aggregation, mitochondrial fragmentation and excess calcium release. *bioRxiv*, p. 2020.2004.2027.053470. doi: 10.1101/2020.04.27.053470

677 Batiuk, M. Y. et al. 2020. Identification of region-specific astrocyte subtypes at single
678 cell resolution. *Nature Communications* 11(1), doi: 10.1038/s41467-019-14198-8

679

680 Bayraktar, O. A. et al. 2020. Astrocyte layers in the mammalian cerebral cortex
681 revealed by a single-cell in situ transcriptomic map. *Nature Neuroscience* 23(4), pp.
682 500-509. doi: 10.1038/s41593-020-0602-1

683

684 Bhaduri, A. et al. 2021. An atlas of cortical arealization identifies dynamic molecular
685 signatures. *Nature* 598(7879), pp. 200-204. doi: 10.1038/s41586-021-03910-8

686

687 Bifsha, P., Balsalobre, A. and Drouin, J. 2017. Specificity of Pitx3-Dependent Gene
688 Regulatory Networks in Subsets of Midbrain Dopamine Neurons. *Molecular*
689 *Neurobiology* 54(7), pp. 4921-4935. doi: 10.1007/s12035-016-0040-y

690

691 Booth, H. D. E., Hirst, W. D. and Wade-Martins, R. 2017. The Role of Astrocyte
692 Dysfunction in Parkinson's Disease Pathogenesis. *Trends in neurosciences* 40(6), pp.
693 358-370. doi: 10.1016/j.tins.2017.04.001

694

695 Booth, H. D. E. et al. 2019. RNA sequencing reveals MMP2 and TGFB1
696 downregulation in LRRK2 G2019S Parkinson's iPSC-derived astrocytes. *Neurobiol*
697 *Dis* 129, pp. 56-66. doi: 10.1016/j.nbd.2019.05.006

698

699 Bradley, R. A. et al. 2019. Regionally specified human pluripotent stem cell-derived
700 astrocytes exhibit different molecular signatures and functional properties.
701 *Development* 146(13), p. dev170910. doi: 10.1242/dev.170910

702

703 Cardo, L. F., Monzón-Sandoval, J., Li, Z., Webber, C. and Li, M. 2023. Single-Cell
704 Transcriptomics and In Vitro Lineage Tracing Reveals Differential Susceptibility of
705 Human iPSC-Derived Midbrain Dopaminergic Neurons in a Cellular Model of
706 Parkinson's Disease. *Cells* 12(24), doi: 10.3390/cells12242860

707

708 Carlson, M. 2019. org.Hs.eg.db: Genome wide annotation for Human. R package
709 version 3.8.2.

710

Chai, H. et al. 2017. Neural Circuit-Specialized Astrocytes: Transcriptomic, Proteomic, Morphological, and Functional Evidence. *Neuron* 95(3), pp. 531-549 e539. doi: 10.1016/j.neuron.2017.06.029

Chandrasekaran, A., Avci, H. X., Leist, M., Kobolák, J. and Dinnyés, A. 2016. Astrocyte Differentiation of Human Pluripotent Stem Cells: New Tools for Neurological Disorder Research. *Frontiers in cellular neuroscience* 10, pp. 215-215. doi: 10.3389/fncel.2016.00215

Chung, N. C. and Storey, J. D. 2015. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* 31(4), pp. 545-554. doi: 10.1093/bioinformatics/btu674

Clough, E. and Barrett, T. 2016. The Gene Expression Omnibus Database. *Methods Mol Biol* 1418, pp. 93-110. doi: 10.1007/978-1-4939-3578-9_5

Crompton, L. A., McComish, S. F., Stathakos, P., Cordero-Llana, O., Lane, J. D. and Caldwell, M. A. 2021. Efficient and Scalable Generation of Human Ventral Midbrain Astrocytes from Human-Induced Pluripotent Stem Cells. *Journal of Visualized Experiments* (176), doi: 10.3791/62095

Crompton, L. A., McComish, S. F., Steward, T. G. J., Whitcomb, D. J., Lane, J. D. and Caldwell, M. A. 2023. Human stem cell-derived ventral midbrain astrocytes exhibit a region-specific secretory profile. *Brain Commun* 5(2), p. fcad114. doi: 10.1093/braincomms/fcad114

de Rus Jacquet, A., Tancredi, J. L., Lemire, A. L., DeSantis, M. C., Li, W. P. and O'Shea, E. K. 2021. The LRRK2 G2019S mutation alters astrocyte-to-neuron communication via extracellular vesicles and induces neuron atrophy in a human iPSC-derived model of Parkinson's disease. *Elife* 10, p. 2020.2007.2002.178574. doi: 10.7554/eLife.73062

Deneen, B., Ho, R., Lukaszewicz, A., Hochstim, C. J., Gronostajski, R. M. and Anderson, D. J. 2006. The Transcription Factor NFIA Controls the Onset of Gliogenesis in the Developing Spinal Cord. *Neuron* 52(6), pp. 953-968. doi: 10.1016/j.neuron.2006.11.019

747

748 di Domenico, A. et al. 2019. Patient-Specific iPSC-Derived Astrocytes Contribute to
749 Non-Cell-Autonomous Neurodegeneration in Parkinson's Disease. *Stem cell reports*
750 12(2), pp. 213-229. doi: 10.1016/j.stemcr.2018.12.011

751

752 Dobin, A. et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1),
753 pp. 15-21. doi: 10.1093/bioinformatics/bts635

754

755 Duan, D., Fu, Y., Paxinos, G. and Watson, C. 2013. Spatiotemporal expression
756 patterns of Pax6 in the brain of embryonic, newborn, and adult mice. *Brain Structure*
757 *and Function* 218(2), pp. 353-372. doi: 10.1007/s00429-012-0397-2

758

759 Endo, F. et al. 2022. Molecular basis of astrocyte diversity and morphology across the
760 CNS in health and disease. *Science* 378(6619), p. eadc9020. doi:
761 doi:10.1126/science.adc9020

762

763 Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. and Kriegstein, A. R. 2021.
764 Single-cell atlas of early human brain development highlights heterogeneity of human
765 neuroepithelial cells and early radial glia. *Nature Neuroscience* 24(4), pp. 584-594.
766 doi: 10.1038/s41593-020-00794-1

767

768 Falk, A. et al. 2012. Capture of neuroepithelial-like stem cells from pluripotent stem
769 cells provides a versatile system for in vitro production of human neurons. *PLOS ONE*
770 7(1), p. e29597. doi: 10.1371/journal.pone.0029597

771

772 Fan, X. et al. 2020. Single-cell transcriptome analysis reveals cell lineage specification
773 in temporal-spatial patterns in human cortical development. *Science Advances* 6(34),
774 p. eaaz2978. doi: doi:10.1126/sciadv.aaz2978

775

776 Finak, G. et al. 2015. MAST: a flexible statistical framework for assessing
777 transcriptional changes and characterizing heterogeneity in single-cell RNA
778 sequencing data. *Genome Biology* 16(1), p. 278. doi: 10.1186/s13059-015-0844-5

779

780 Gabay, L., Lowell, S., Rubin, L. L. and Anderson, D. J. 2003. Deregulation of
781 dorsoventral patterning by FGF confers trilineage differentiation capacity on CNS stem
782 cells in vitro. *Neuron* 40(3), pp. 485-499. doi: 10.1016/s0896-6273(03)00637-8

783

784 Hedegaard, A., Monzón-Sandoval, J., Newey, S. E., Whiteley, E. S., Webber, C. and
785 Akerman, C. J. 2020. Pro-maturational Effects of Human iPSC-Derived Cortical
786 Astrocytes upon iPSC-Derived Cortical Neurons. *Stem cell reports* 15(1), pp. 38-51.
787 doi: https://doi.org/10.1016/j.stemcr.2020.05.003

788

789 Holmqvist, S. et al. 2015. Generation of human pluripotent stem cell reporter lines for
790 the isolation of and reporting on astrocytes generated from ventral midbrain and
791 ventral spinal cord neural progenitors. *Stem Cell Research* 15(1), pp. 203-220. doi:
792 https://doi.org/10.1016/j.scr.2015.05.014

793

794 Houweling, A. C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A. F. M., Rüther,
795 U. and Christoffels, V. M. 2001. Gene and cluster-specific expression of the Iroquois
796 family members during mouse development. *Mechanisms of Development* 107(1), pp.
797 169-174. doi: https://doi.org/10.1016/S0925-4773(01)00451-8

798

799 Itoh, N. et al. 2018. Cell-specific and region-specific transcriptomics in the multiple
800 sclerosis model: Focus on astrocytes. *Proceedings of the National Academy of*
801 *Sciences* 115(2), pp. E302-E309. doi: doi:10.1073/pnas.1716032115

802

803 Jaeger, I. et al. 2011. Temporally controlled modulation of FGF/ERK signaling directs
804 midbrain dopaminergic neural progenitor fate in mouse and human pluripotent stem
805 cells. *Development* 138(20), pp. 4363-4374. doi: 10.1242/dev.066746

806

807 Jain, M., Armstrong, R. J., Tyers, P., Barker, R. A. and Rosser, A. E. 2003. GABAergic
808 immunoreactivity is predominant in neurons derived from expanded human neural
809 precursor cells in vitro. *Exp Neurol* 182(1), pp. 113-123. doi: 10.1016/s0014-
810 4886(03)00055-4

811

812 Kamath, T. et al. 2022. Single-cell genomic profiling of human dopamine neurons
813 identifies a population that selectively degenerates in Parkinson's disease. *Nature*
814 *Neuroscience* 25(5), pp. 588-595. doi: 10.1038/s41593-022-01061-1

815

816   Koch, P., Opitz, T., Steinbeck, J. A., Ladewig, J. and Brüstle, O. 2009. A rosette-type,
817   self-renewing human ES cell-derived neural stem cell with potential for in vitro
818   instruction and synaptic integration. *Proc Natl Acad Sci U S A* 106(9), pp. 3225-3230.
819   doi: 10.1073/pnas.0808387106

820

821   Kostuk, E. W., Cai, J. and Iacovitti, L. 2019. Subregional differences in astrocytes
822   underlie selective neurodegeneration or protection in Parkinson's disease models in
823   culture. *Glia* 67(8), pp. 1542-1557. doi: 10.1002/glia.23627

824

825   Krencik, R., Weick, J. P., Liu, Y., Zhang, Z.-J. and Zhang, S.-C. 2011. Specification of
826   transplantable astroglial subtypes from human pluripotent stem cells. *Nature*
827   *Biotechnology* 29, p. 528. doi: 10.1038/nbt.1877

828   https://www.nature.com/articles/nbt.1877#supplementary-information

829

830   La Manno, G. et al. 2016. Molecular Diversity of Midbrain Development in Mouse,
831   Human, and Stem Cells. *Cell* 167(2), pp. 566-580.e519. doi:
832   10.1016/j.cell.2016.09.027

833

834   Li, Y. et al. 2023. Spatiotemporal transcriptome atlas reveals the regional specification
835   of the developing human brain. *Cell* 186(26), pp. 5892-5909.e5822. doi:
836   https://doi.org/10.1016/j.cell.2023.11.016

837

838   Liao, Y., Smyth, G. K. and Shi, W. 2013. The Subread aligner: fast, accurate and
839   scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41(10), p. e108. doi:
840   10.1093/nar/gkt214

841

842   Lin, Y.-T. et al. 2018. APOE4 Causes Widespread Molecular and Cellular Alterations
843   Associated with Alzheimer's Disease Phenotypes in Human iPSC-Derived Brain Cell
844   Types.   *Neuron*   98(6),   pp.   1141-1154.e1147.   doi:
845   https://doi.org/10.1016/j.neuron.2018.05.008

846

847   Liu, Y. et al. 2004. CD44 expression identifies astrocyte-restricted precursor cells.
848   *Developmental   Biology*   276(1),   pp.   31-46.   doi:
849   https://doi.org/10.1016/j.ydbio.2004.08.018

850

851 Lozzi, B., Huang, T. W., Sardar, D., Huang, A. Y. S. and Deneen, B. 2020. Regionally
852 Distinct Astrocytes Display Unique Transcription Factor Profiles in the Adult Brain.
853 *Frontiers in Neuroscience* 14, doi: 10.3389/fnins.2020.00061

854

855 Makarava, N., Chang, J. C.-Y., Kushwaha, R. and Baskakov, I. V. 2019. Region-
856 Specific Response of Astrocytes to Prion Infection. *Frontiers in Neuroscience* 13, doi:
857 10.3389/fnins.2019.01048

858

859 Matsunaga, E., Araki, I. and Nakamura, H. 2001. Role of Pax3/7 in the tectum
860 regionalization. *Development* 128(20), pp. 4069-4077. doi: 10.1242/dev.128.20.4069

861

862 McCarthy, D. J., Campbell, K. R., Lun, A. T. L. and Wills, Q. F. 2017. Scater: pre-
863 processing, quality control, normalization and visualization of single-cell RNA-seq data
864 in R. *Bioinformatics* 33(8), pp. 1179-1186. doi: 10.1093/bioinformatics/btw777

865

866 Molofsky, A. V. et al. 2012. Astrocytes and disease: a neurodevelopmental
867 perspective. *Genes & Development* 26(9), pp. 891-907. doi: 10.1101/gad.188326.112

868

869 Morel, L. et al. 2017. Molecular and Functional Properties of Regional Astrocytes in
870 the Adult Brain. *J Neurosci* 37(36), pp. 8706-8717. doi: 10.1523/JNEUROSCI.3956-
871 16.2017

872

873 Nolbrant, S., Heuer, A., Parmar, M. and Kirkeby, A. 2017. Generation of high-purity
874 human ventral midbrain dopaminergic progenitors for in vitro maturation and
875 intracerebral transplantation. *Nature Protocols* 12(9), pp. 1962-1979. doi:
876 10.1038/nprot.2017.078

877

878 Oberheim, N. A. et al. 2009. Uniquely hominid features of adult human astrocytes. *J
879 Neurosci* 29(10), pp. 3276-3287. doi: 10.1523/jneurosci.4707-08.2009

880

881 Peteri, U.-K., Pitkonen, J., Utami, K. H., Paavola, J., Roybon, L., Pouladi, M. A. and
882 Castrén, M. L. 2021. Generation of the Human Pluripotent Stem-Cell-Derived
883 Astrocyte Model with Forebrain Identity. *Brain sciences* 11(2), p. 209. doi:
884 10.3390/brainsci11020209

885

886 Phatnani, H. and Maniatis, T. 2015. Astrocytes in neurodegenerative disease. *Cold*
887 *Spring Harbor perspectives in biology* 7(6), p. a020628. doi:
888 10.1101/cshperspect.a020628

889

890 Polioudakis, D. et al. 2019. A Single-Cell Transcriptomic Atlas of Human Neocortical
891 Development during Mid-gestation. *Neuron* 103(5), pp. 785-801.e788. doi:
892 10.1016/j.neuron.2019.06.011

893

894 Roybon, L. et al. 2013. Human stem cell-derived spinal cord astrocytes with defined
895 mature or reactive phenotypes. *Cell Rep* 4(5), pp. 1035-1048. doi:
896 10.1016/j.celrep.2013.06.021

897

898 Schindelin, J. et al. 2012. Fiji: an open-source platform for biological-image analysis.
899 *Nature Methods* 9(7), pp. 676-682. doi: 10.1038/nmeth.2019

900

901 Schlicker, A., Domingues, F. S., Rahnenführer, J. and Lengauer, T. 2006. A new
902 measure for functional similarity of gene products based on Gene Ontology. *BMC*
903 *Bioinformatics* 7(1), p. 302. doi: 10.1186/1471-2105-7-302

904

905 Schober, A. L., Wicki-Stordeur, L. E., Murai, K. K. and Swayne, L. A. 2022.
906 Foundations and implications of astrocyte heterogeneity during brain development
907 and disease. *Trends in neurosciences* 45(9), pp. 692-703. doi:
908 10.1016/j.tins.2022.06.009

909

910 Serio, A. et al. 2013. Astrocyte pathology and the absence of non-cell autonomy in an
911 induced pluripotent stem cell model of TDP-43 proteinopathy. *Proceedings of the*
912 *National Academy of Sciences of the United States of America* 110(12), pp. 4697-
913 4702. doi: 10.1073/pnas.1300398110

914

915 Sloan, S. A. et al. 2017. Human Astrocyte Maturation Captured in 3D Cerebral Cortical
916 Spheroids Derived from Pluripotent Stem Cells. *Neuron* 95(4), pp. 779-790 e776. doi:
917 10.1016/j.neuron.2017.07.035

918

919 Sonninen, T.-M. et al. 2020. Metabolic alterations in Parkinson's disease astrocytes.
920 *Scientific Reports* 10(1), p. 14474. doi: 10.1038/s41598-020-71329-8

921

922 Stirling, D. R., Swain-Bowden, M. J., Lucas, A. M., Carpenter, A. E., Cimini, B. A. and
923 Goodman, A. 2021. CellProfiler 4: improvements in speed, utility and usability. *BMC*
924 *Bioinformatics* 22(1), p. 433. doi: 10.1186/s12859-021-04344-9

925

926 Stolt, C. C., Lommes, P., Sock, E., Chaboissier, M.-C., Schedl, A. and Wegner, M.
927 2003. The Sox9 transcription factor determines glial fate choice in the developing
928 spinal cord. *Genes & Development* 17(13), pp. 1677-1689. doi: 10.1101/gad.259003

929

930 Strelau, J. et al. 2000. Growth/differentiation factor-15/macrophage inhibitory cytokine-
931 1 is a novel trophic factor for midbrain dopaminergic neurons in vivo. *J Neurosci*
932 20(23), pp. 8597-8603. doi: 10.1523/jneurosci.20-23-08597.2000

933

934 Stuart, T. et al. 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177(7), pp.
935 1888-1902.e1821. doi: 10.1016/j.cell.2019.05.031

936

937 Sun, Y. et al. 2008. Long-term tripotent differentiation capacity of human neural stem
938 (NS) cells in adherent culture. *Mol Cell Neurosci* 38(2), pp. 245-258. doi:
939 10.1016/j.mcn.2008.02.014

940

941 Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. 2011. REVIGO Summarizes and
942 Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* 6(7), p. e21800. doi:
943 10.1371/journal.pone.0021800

944

945 Takata, N. and Hirase, H. 2008. Cortical Layer 1 and Layer 2/3 Astrocytes Exhibit
946 Distinct Calcium Dynamics In Vivo. *PLOS ONE* 3(6), p. e2525. doi:
947 10.1371/journal.pone.0002525

948

949 Tcw, J. et al. 2017. An Efficient Platform for Astrocyte Differentiation from Human
950 Induced Pluripotent Stem Cells. *Stem cell reports* 9(2), pp. 600-614. doi:
951 10.1016/j.stemcr.2017.06.018

952

953   Team, R. C. 2023. : A Language and Environment for Statistical Computing. Vienna,
954   Austria: R Foundation for Statistical Computing, .

955

956   Verkhratsky, A. and Nedergaard, M. 2018. Physiology of astroglia. *Physiological*
957   *Reviews* 98(1), pp. 239-389. doi: 10.1152/physrev.00042.2016

958

959   Xin, W., Schuebel, K. E., Jair, K.-w., Cimbro, R., De Biase, L. M., Goldman, D. and
960   Bonci, A. 2019. Ventral midbrain astrocytes display unique physiological features and
961   sensitivity to dopamine D2 receptor signaling. *Neuropsychopharmacology* 44(2), pp.
962   344-355. doi: 10.1038/s41386-018-0151-4

963

964   Yun, W. et al. 2019. Generation of Anterior Hindbrain-Specific, Glial-Restricted
965   Progenitor-Like Cells from Human Pluripotent Stem Cells. *Stem Cells and*
966   *Development* 28(10), pp. 633-648. doi: 10.1089/scd.2019.0033

967

968   Zhong, S. et al. 2018. A single-cell RNA-seq survey of the developmental landscape
969   of the human prefrontal cortex. *Nature* 555(7697), pp. 524-528. doi:
970   10.1038/nature25980

971

972   Zhou, S. et al. 2016. Neurosphere Based Differentiation of Human iPSC Improves
973   Astrocyte Differentiation. *Stem Cells Int* 2016, p. 4937689. doi: 10.1155/2016/4937689

974

975   Zhu, Q., Shah, S., Dries, R., Cai, L. and Yuan, G. C. 2018. Identification of spatially
976   associated subpopulations by combining scRNAseq and sequential fluorescence in
977   situ hybridization data. *Nat Biotechnol* 36(12), pp. 1183-1190. doi: 10.1038/nbt.4260

978

979