Cell Reports, Volume *32*

Supplemental Information

Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans

Nicola Thrupp, Carlo Sala Frigerio, Leen Wolfs, Nathan G. Skene, Nicola Fattorelli, Suresh Poovathingal, Yannick Fourne, Paul M. Matthews, Tom Theys, Renzo Mancuso, Bart de Strooper, and Mark Fiers

Figure S1

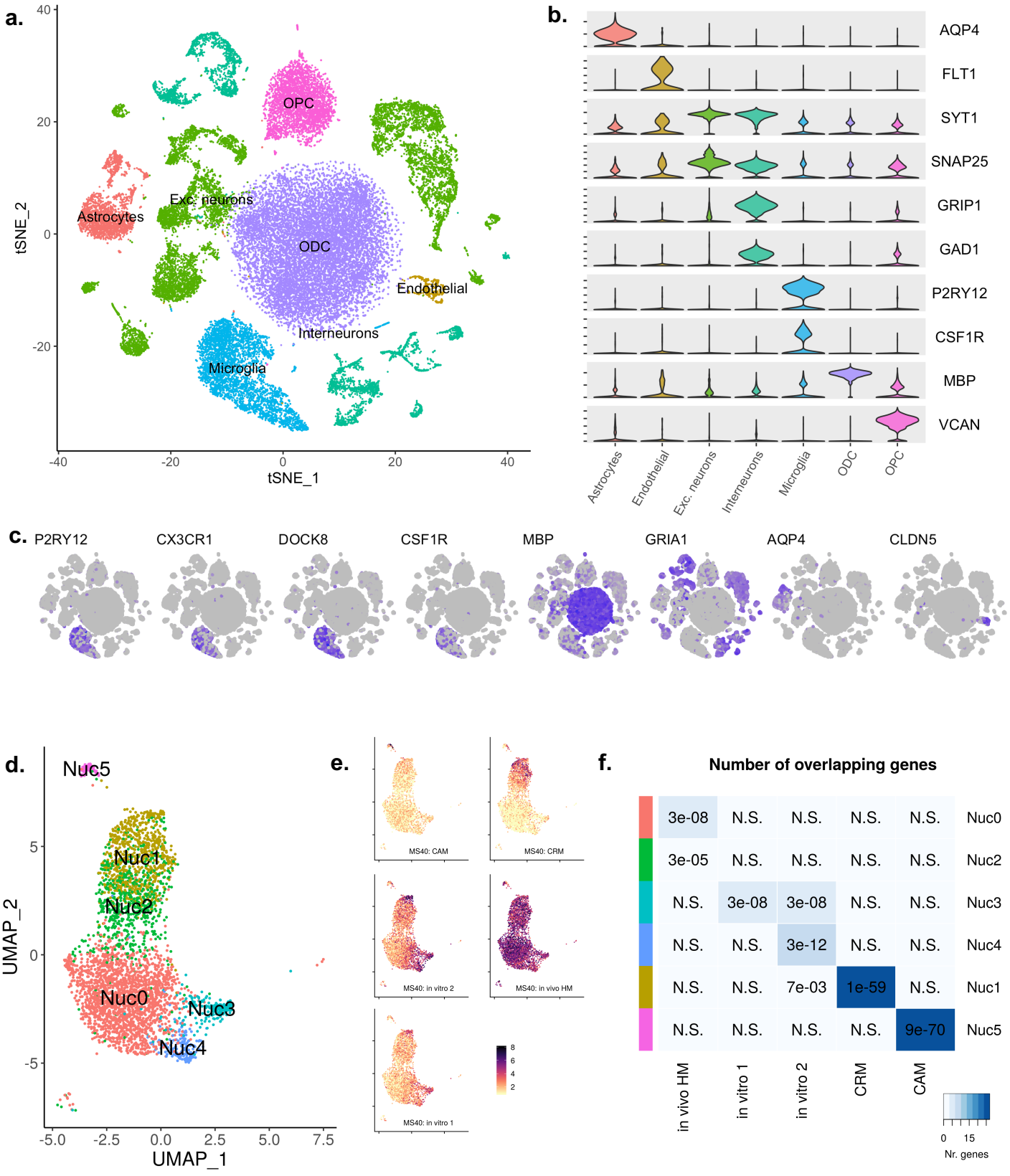
a.
tSNE_2
tSNE_1
OPC
Exc. neurons
Astrocytes
ODC
Endothelial
Interneurons
Microglia
b.
AQP4
FLT1
SYT1
SNAP25
GRIP1
GAD1
P2RY12
CSF1R
MBP
VCAN
Astrocytes
Endothelial
Exc. neurons
Interneurons
Microglia
ODC
OPC
c.
P2RY12
CX3CR1
DOCK8
CSF1R
MBP
GRIA1
AQP4
CLDN5
d.
Nuc5
Nuc1
Nuc2
Nuc0
Nuc3
Nuc4
UMAP_2
UMAP_1
e.
MS40: CAM
MS40: CRM
MS40: in vitro 2
MS40: in vivo HM
MS40: in vitro 1
f.
Number of overlapping genes
3e-08 N.S. N.S. N.S. N.S. Nuc0
3e-05 N.S. N.S. N.S. N.S. Nuc2
N.S. 3e-08 3e-08 N.S. N.S. Nuc3
N.S. N.S. 3e-12 N.S. N.S. Nuc4
N.S. N.S. 7e-03 1e-59 N.S. Nuc1
N.S. N.S. N.S. N.S. 9e-70 Nuc5
in vivo HM
in vitro 1
in vitro 2
CRM
CAM
Nr. genes

Fig. S1: Clustering of single nuclei from human tissue. Related to STAR Methods.

a. tSNE plot of 37,060 nuclei from cortical tissue of 4 neurosurgical patients, coloured according to cell type. OPC, oligodendrocyte precursor cells ; ODC, oligodendrocytes ; Exc. neurons, excitatory neurons. **b.** Violin plots show selected markers of the different cell types (data is normalised for count depth and log-transformed using the natural log). **c.** tSNE plot of all nuclei, highlighting microglial and other celltype marker genes. Grey = 0 expression ; blue = high expression. **d.** UMAP plot of 3,721 microglial nuclei from cortical tissue of 4 neurosurgical patients, coloured according to cluster number, after *in silico* extraction of microglia (based on markers such as *P2RY12*) and reclustering. **e.** Module scores for gene sets extracted from the original Mancuso *et al.* single cell microglia study (Mancuso et al., 2019). The top 40 genes according to log fold change were selected for each gene set. **f.** Overlap of top 40 marker genes from cellular clusters on the horizontal axis (Mancuso *et al.,* 2019) and nuclear clusters on the vertical axis. The blue scale represents the number of genes in common, numbers represent p_{adj} values. Vertical coloured bars correspond to the clusters shown in c). N.S., not significant ($p_{adj} > 0.05$). MS40, Module Score of top 40 gene markers ; CAM, macrophages ; CRM, cytokine response microglia ; in vitro 1, activation-like module (similar to *in vitro* macrophages) ; in vitro 2, activation-like module (similar to *in vitro* monocytes) ; in vivo HM, homeostatic ; Nuc, Nuclear clusters. Cluster markers are provided in Supplementary Table 1.

Figure S2

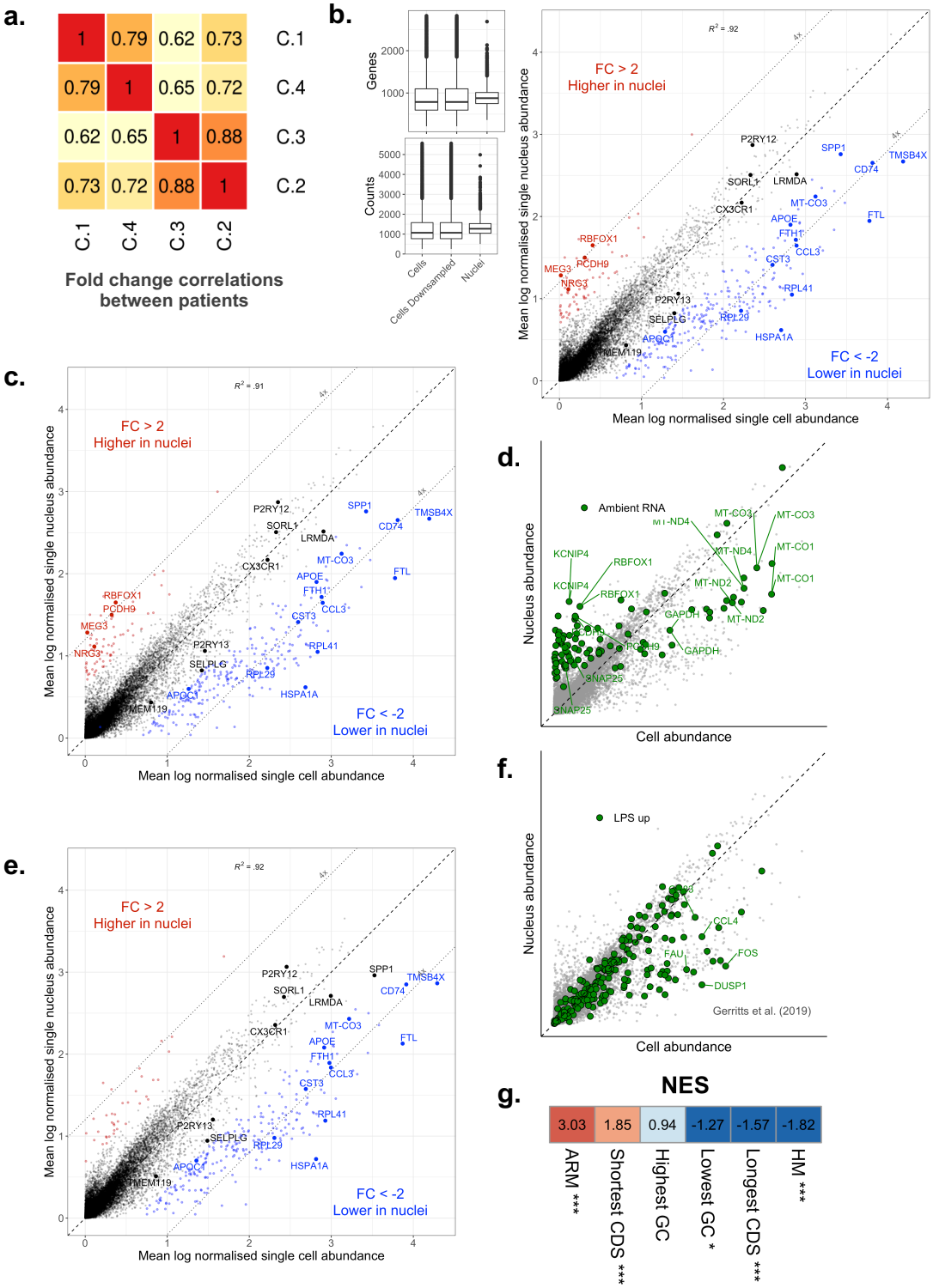
a.
1 0.79 0.62 0.73 C.1
0.79 1 0.65 0.72 C.4
0.62 0.65 1 0.88 C.3
0.73 0.72 0.88 1 C.2
C.1 C.4 C.3 C.2
Fold change correlations between patients
b.
Genes
Counts
Cells
Cells Downsampled
Nuclei
FC > 2
Higher in nuclei
FC < -2
Lower in nuclei
Mean log normalised single nucleus abundance
Mean log normalised single cell abundance
c.
d.
Ambient RNA
Nucleus abundance
Cell abundance
e.
f.
LPS up
Gerritts et al. (2019)
g.
NES
3.03 1.85 0.94 -1.27 -1.57 -1.82
ARM ***
Shortest CDS ***
Highest GC
Lowest GC *
Longest CDS ***
HM ***

Fig. S2: Gene abundance in single microglial cells versus single microglial nuclei of human cortical tissue. Related to Figure 1 and Figure 2.

a. Correlation matrix of gene abundance fold changes (single cell vs single nucleus abundance) between the four patients. **b**. (L) Downsampling of reads: boxplots for numbers of reads and numbers of genes for single cells before downsampling, single cells after downsampling, and single nuclei. Boxplots show median, with 25% and 75% quantiles. (R) Scatterplot of mean gene abundance in cells against mean gene abundance in nuclei (as in Fig. 1a) after downsampling of reads in cells. Points in red represent genes with significantly higher abundance in nuclei, while those in blue are significantly less abundant in nuclei ($p_{adj} < 0.05$, fold change > |2|). **c.** Downsampling of cells. Scatterplot of mean gene abundance in cells against mean gene abundance in nuclei (as in Fig. 1a) after downsampling of cells. Points in red represent genes with significantly higher abundance in nuclei, while those in blue are significantly less abundant in nuclei ($p_{adj} < 0.05$, fold change > |2|). **d.** Scatter plot, as in Fig. 1a) showing the ambient mRNA in green (the same dataset was used in Fig. 2d). Ambient RNA is defined as the 300 most abundant genes in the 700 nuclei with the lowest total number of reads. **e.** Scatterplot as in Fig. 1a., with ambient genes removed before normalization. Mean normalised gene abundance in cells (x axis) and nuclei. (y axis). Red: genes with significantly higher abundance in nuclei ($p_{adj} < 0.05$, fold change > 2). Blue: genes that are significantly less abundance in nuclei ($p_{adj} < 0.05$, fold change < -2). **f.** Scatterplot as in Fig. 1a), highlighting in green genes that are upregulated during LPS stimulation in mice (Gerrits et al., 2019). A regression line for the highlighted genes is shown in green (slope = 0.78). **g.** Gene Set Enrichment Analysis (GSEA) of gene sets related to cellular location and gene coding sequence length (CDS). Background genes were ranked according to log fold change of ARM genes versus homeostatic genes (Sala Frigerio et al., 2019). Red: higher Normalised Enrichment Score (NES), i.e. more activation genes ; blue: negative NES scores (more homeostatic genes). For figures b-f, gene abundances were normalized to read depth per cell, scaled by 10,000 and log-transformed using the natural log. *MALAT1* (which had normalized abundance levels of 6.0 and 6.9 respectively in cells and nuclei) has been removed for visualisation purposes. * represents significance ($p_{adj} < 0.05$). *** represents significance ($p_{adj} < 0.0005$). FC, fold change; CDS, coding sequence; R2, correlation coefficient.

Figure S3

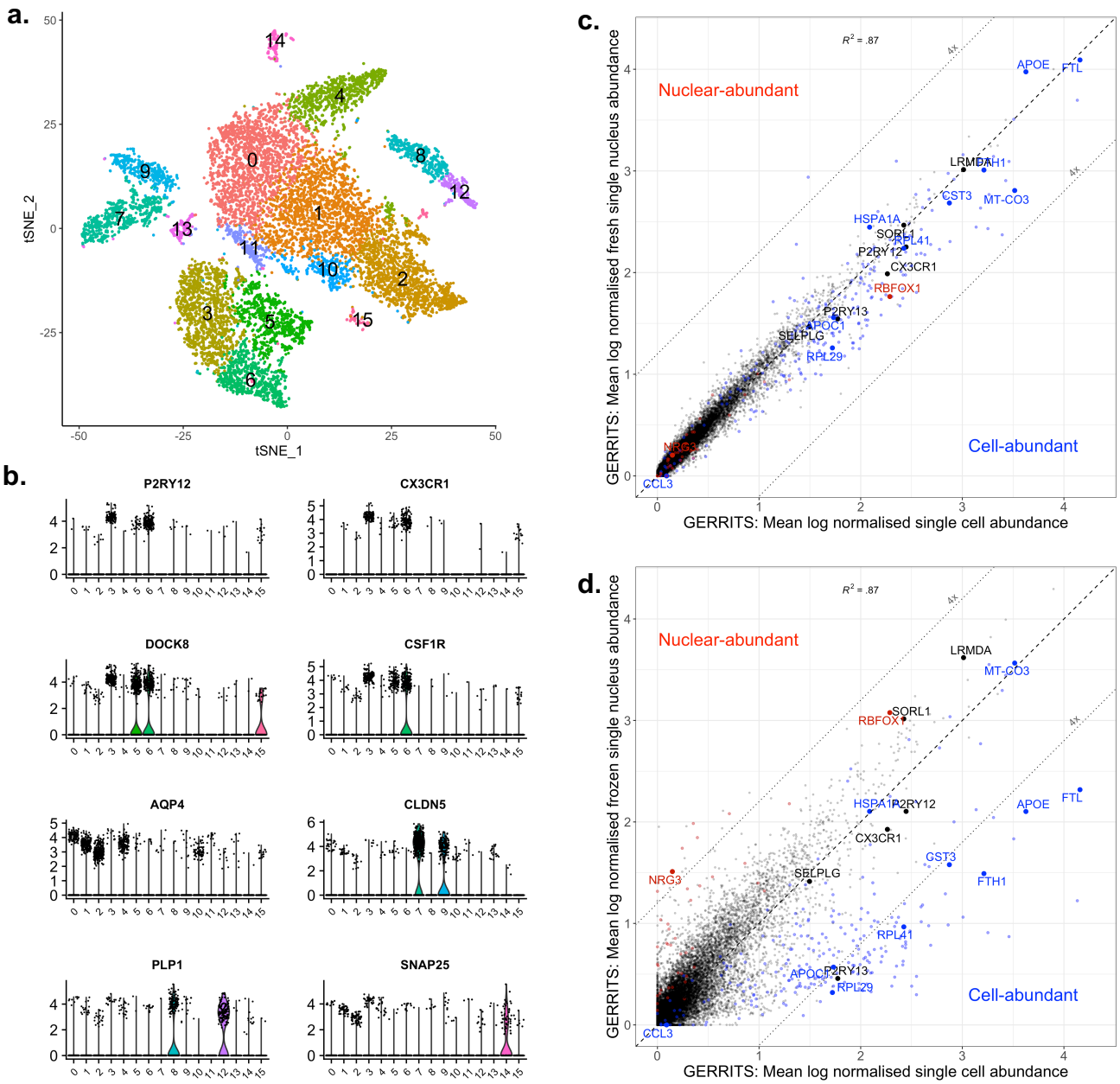
a.
tSNE_1
tSNE_2
b.
P2RY12
CX3CR1
DOCK8
CSF1R
AQP4
CLDN5
PLP1
SNAP25
c.
R^2 = .87
Nuclear-abundant
Cell-abundant
GERRITS: Mean log normalised fresh single nucleus abundance
GERRITS: Mean log normalised single cell abundance
d.
R^2 = .87
Nuclear-abundant
Cell-abundant
GERRITS: Mean log normalised frozen single nucleus abundance
GERRITS: Mean log normalised single cell abundance

Fig. S3: Consideration of previous studies. Related to Figure 2.
a. UMAP projection of frozen nuclei from the Gerrits *et al*. study of human nuclei. The projection is coloured according to clusters identified by Louvain clustering. **b**. Violin plots of know markers of microglia (*P2RY12, CX3CR1, DOCK8, CSF1R*), oligodendrocytes (*PLP1*), neurons (*SNAP25*), astrocytes (*AQP4*) and endothelial cells (*CLDN5*). Marker expression levels are shown for each cluster identified in Fig. S3a. **c,d.** Mean log-normalised cell abundance (x axis) against, **c.** mean log-normalised fresh single nucleus abundance, and **d.** mean log-normalised frozen single nucleus abundance, of human microglia captured by Gerrits *et al*. Red: genes with significantly higher abundance in nuclei in our data ($p_{adj} < 0.05$, fold change > 2). Blue: genes that are significantly less abundance in nuclei ($p_{adj} < 0.05$, fold change < -2) in our data. The black dashed line represents no fold change; the grey dotted lines represent a 4-fold differences between cells and nuclei. Gene abundances were normalized to read depth per cell, scaled by 10,000 and log-transformed using the natural log. FC, fold change.