

# Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry

Chotima Böttcher<sup>1,14\*</sup>, Stephan Schlickeiser<sup>2,14</sup>, Marjolein A. M. Sneuboer<sup>3,14</sup>, Desiree Kunkel<sup>2</sup>, Annik Knop<sup>1</sup>, Evdokia Paza<sup>1</sup>, Pawel Fidzinski<sup>4</sup>, Larissa Kraus<sup>4,5</sup>, Gijsje J. L. Snijders<sup>3</sup>, René S Kahn<sup>6</sup>, Axel R Schulz<sup>7</sup>, Henrik E Mei<sup>7</sup>, NBB-Psy<sup>8</sup>, Elly M. Hol<sup>3,9</sup>, Britta Siegmund<sup>10</sup>, Rainer Glaben<sup>10</sup>, Eike J Spruth<sup>1,11</sup>, Lot D de Witte<sup>3,6,15</sup> and Josef Priller<sup>1,5,11,12,13,15\*</sup>

**Microglia, the specialized innate immune cells of the CNS, play crucial roles in neural development and function. Different phenotypes and functions have been ascribed to rodent microglia, but little is known about human microglia (huMG) heterogeneity. Difficulties in procuring huMG and their susceptibility to cryopreservation damage have limited large-scale studies. Here we applied multiplexed mass cytometry for a comprehensive characterization of postmortem huMG (10<sup>3</sup> – 10<sup>4</sup> cells). We determined expression levels of 57 markers on huMG isolated from up to five different brain regions of nine donors. We identified the phenotypic signature of huMG, which was distinct from peripheral myeloid cells but was comparable to fresh huMG. We detected microglia regional heterogeneity using a hybrid workflow combining Cytobank and R/Bioconductor for multidimensional data analysis. Together, these methodologies allowed us to perform high-dimensional, large-scale immunophenotyping of huMG at the single-cell level, which facilitates their unambiguous profiling in health and disease.**

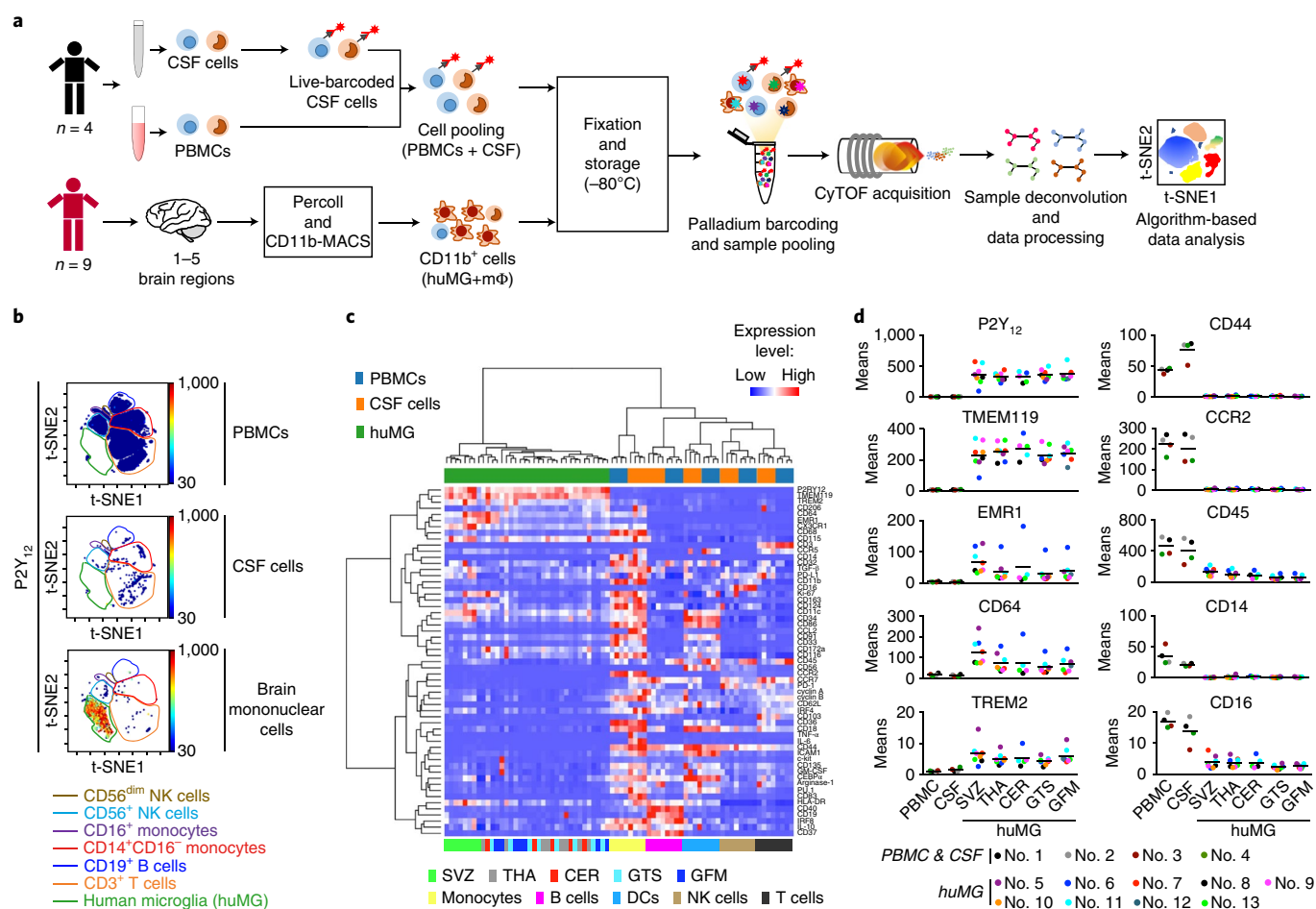
Microglia are resident innate immune cells of the CNS, and have an important role in maintaining CNS integrity and function<sup>1</sup>. They are involved in removing apoptotic neurons, refining synaptic connectivity and providing trophic support for memory and motor learning<sup>2,3</sup>. In addition, microglia play a role in the development and progression of neurological and psychiatric disorders, including Alzheimer's disease, amyotrophic lateral sclerosis, schizophrenic psychoses and mood disorders<sup>4–7</sup>. Thus, the identification of mechanisms that regulate microglial homeostasis and function may provide the means to manipulate these cells for therapeutic purposes. Over the last two decades, microglial ontogeny, phenotypic heterogeneity and responses to CNS pathology have been extensively studied in rodents<sup>8–15</sup>. However, much less is known about human microglia (huMG). Comparative studies of the transcriptional network revealed overall similarity in the transcriptomic landscapes of human and mouse microglia<sup>16,17</sup>. Nonetheless, these studies also demonstrated species-specific patterns of gene expression, and differences in the responses of human and murine microglia to aging<sup>16,17</sup>. The heterogeneity of microglia that has been described for the mouse brain<sup>7,14</sup> has so far not been replicated in humans. It is also unclear how human and mouse microglia compare on a phenotypic level.

So far, the phenotypic characterization of huMG has mainly relied on immunohistochemical analysis of postmortem brain tissue,

fluorescence flow cytometric analysis of isolated microglial cells or in vitro culture of huMG<sup>18–21</sup>. Among the limitations of these approaches are the phenotypic changes induced by cell culture<sup>16,19</sup>, the high autofluorescent background of postmortem tissue and the restrictions in the number of markers that can be simultaneously investigated in one measurement (commonly less than 20). In addition, due to the lack of a validated protocol for cryopreservation, flow cytometric analysis of huMG has been restricted to immediate measurements of acutely isolated cells, which can result in batch effects and contribute to erroneous interpretations of data.

In this study, we developed a method to deep profile the immune phenotype of small samples of huMG at the single-cell level. The protocol allows for simultaneous measurement of multiple samples from different donors and brain regions, and at the same time for comparison with cells from other compartments (for example, cerebrospinal fluid (CSF) and peripheral blood). We isolated huMG from different regions of postmortem brain tissues and from fresh brain biopsies following an established protocol. Peripheral immune cells were freshly isolated from blood and CSF. The isolated cells (both huMG and peripheral immune cells) were fixed and stored using a protocol for long-term cryopreservation. Subsequently, cryopreserved huMG and peripheral immune cells were simultaneously profiled by multiplexed mass cytometry (CyTOF) using

<sup>1</sup>Department of Neuropsychiatry and Laboratory of Molecular Psychiatry, Charité—Universitätsmedizin Berlin, Berlin, Germany. <sup>2</sup>Berlin-Brandenburg Center for Regenerative Therapies, Berlin, Germany. <sup>3</sup>Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands. <sup>4</sup>Epilepsy-Center Berlin-Brandenburg, Department of Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany. <sup>5</sup>Berlin Institute of Health, Berlin, Germany. <sup>6</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>7</sup>German Rheumatism Research Center, Berlin, Germany. <sup>8</sup>Netherlands Brain Bank, Amsterdam, the Netherlands. <sup>9</sup>Department of Neuroimmunology, Netherlands Institute for Neuroscience, An Institute of the Royal Academy of Arts and Sciences, Amsterdam, the Netherlands. <sup>10</sup>Medical Department for Gastroenterology, Division of Gastroenterology, Infectiology and Rheumatology, Charité – Universitätsmedizin Berlin, Berlin, Germany. <sup>11</sup>DZNE, Berlin, Germany. <sup>12</sup>Cluster of Excellence NeuroCure, Berlin, Germany. <sup>13</sup>University of Edinburgh and UK Dementia Research Institute, Edinburgh, UK. <sup>14</sup>These authors contributed equally: Chotima Böttcher, Stephan Schlickeiser, Marjolein A. M. Sneuboer. <sup>15</sup>These authors jointly supervised this work: Lot D. de Witte, Josef Priller. \*e-mail: [chotima.boettcher@charite.de](mailto:chotima.boettcher@charite.de); [josef.priller@charite.de](mailto:josef.priller@charite.de)



**Fig. 1 | Simultaneous high-dimensional immune phenotyping of human microglia, and mononuclear cells from blood and CSF. a**, Schematic representation of sample processing and CyTOF measurement. Blood and CSF were collected from the same individuals. Human microglia (huMG) were isolated from subventricular zone (SVZ), thalamus (THA), cerebellum (CER), temporal lobe (GTS), and frontal lobe (GFM) from nine biologically independent donors. One final barcoded and pooled sample consists of 18 huMG samples and 2 PBMC-CSF sample pools. Mixed samples were equally divided and stained with two panels of metal-conjugated antibodies and acquired on the CyTOF instrument. **b**, Representative two-dimensional projections of single-cell data generated by t-SNE of biologically independent samples: PBMCs ( $n=4$ ), CSF cells ( $n=4$ ) and brain mononuclear cells ( $n=36$ ). Each dot represents one cell. The color spectrum represents expression of P2Y<sub>12</sub> (red, high expression; blue, no expression). P2Y<sub>12</sub><sup>+</sup> cells were gated as huMG (green) and P2Y<sub>12</sub><sup>-</sup> cells were gated as different circulating immune cells. **c**, Heat map and cluster analysis of all samples on the basis of the mean expression of 57 markers (Panel A plus Panel B). Similarities between PBMCs (blue), CSF cells (orange) and huMG (green), as well as the similarities between huMG from different brain regions (SVZ, bright green; THA, gray; CER, red; GTS, bright blue; GFM, dark blue) samples and expression levels are indicated by dendrograms. Heat colors of expression levels have been scaled for each marker individually (to the 1st and 5th quintiles) (red, high expression; blue, low expression). **d**, Mean signal intensity levels of P2Y<sub>12</sub>, TMEM119, EMR1 (F4/80), CD64, TREM2, CD44, CCR2, CD45, CD14 and CD16 staining in PBMCs, CSF cells and huMG from different brain regions (black lines show mean values of the datasets).

barcoding technology. These powerful methodologies allowed accurate and unbiased analysis of an unprecedented number of markers at the single-cell level. Our results reveal a unique phenotypic signature of huMG that distinguishes them from other mononuclear cells in the CSF and peripheral blood. Using a hybrid Cytobank- and R/Bioconductor-based data processing and analysis workflow, we provide evidence for the heterogeneity of microglia in the human brain.

## Results

**Mass cytometric analysis of cryopreserved human postmortem brain microglia.** To compare the different phenotypes of CNS-resident microglia and peripheral immune cells, we simultaneously profiled peripheral blood mononuclear cells (PBMCs), immune cells from the CSF and huMG in the same run. The experiment is outlined in Fig. 1a. In summary, huMG were isolated from postmortem brain tissue of different brain regions as described previously<sup>19</sup>. Isolated huMG were then cryopreserved at  $-80^{\circ}\text{C}$  using

paraformaldehyde-containing stabilizing buffer<sup>22</sup>. Of note is that classical cryopreservation using DMSO failed to cryopreserve the isolated huMG. Up to  $3 \times 10^4$  live CSF-derived cells were CD45-barcode with <sup>89</sup>Y-CD45 antibody (<sup>89</sup>Y-CD45<sup>+</sup>)<sup>23</sup> and subsequently pooled with PBMCs (<sup>89</sup>Y-CD45<sup>-</sup>) from the same individual. The CSF-PBMC pooled samples were then cryopreserved using the same protocol that was applied to the huMG. To minimize the run-to-run variation and to facilitate the comparison of cellular phenotypes from different compartments and individuals, we thawed huMG and CSF-PBMC samples and performed intracellular mass-tag barcoding using palladium (Pd) isotopes. Each sample pool consisted of two CSF-PBMC sample pairs (two individuals with CSF and PBMC each) and 18 huMG samples (from up to five brain regions of 4–5 donors) (Supplementary Table 1). The pooled samples were split equally and stained with two different antibody panels (35 antibodies/panel) (Supplementary Tables 2 and 3). Panel A was designed to encompass the major circulating immune cell subsets (T and B cells,

monocytes, natural killer (NK) cells) and microglia using proliferation markers, activity-related markers, chemokine receptors and cell subset markers, including P2Y<sub>12</sub>, IRF4, IRF8, CD45, CD3, CD62L, CD19, HLA-DR, CD56, cyclin A and B1 and Ki-67. Panel B was designed to analyze the phenotypes of huMG and the innate immune cell subsets using 35 antibodies, including TMEM119, CD172a, CD279 (PD-1), CD274 (PD-L1), arginase-1, CCR7, CD44, CD18 and CD32. Finally, barcoded and pooled samples were simultaneously acquired on a CyTOF instrument.

To capture and visualize all mononuclear cell subpopulations in a single two-dimensional (2D) map, we first performed an unsupervised high-dimensional data analysis using the t-distributed stochastic linear embedding (t-SNE) algorithm<sup>24,25</sup> on the commercially available analysis platform Cytobank ([www.cytobank.org](http://www.cytobank.org)) (Fig. 1b). The t-SNE maps showed a unique and distinct cluster of the huMG samples (Fig. 1b, green gate). This cluster expressed for the microglial marker, P2Y<sub>12</sub>, and on the basis of previous work<sup>12,20</sup> was used to identify human microglia (huMG). Notably, P2Y<sub>12</sub>-negative cells detected in the brain samples (Fig. 1b) showed similar t-SNE coordination in clusters that overlapped with circulating immune cells in the peripheral blood and CSF (Fig. 1b). A unique and distinct cluster of huMG was also identified without the markers traditionally used to identify microglia (for example, CD11b and CD45) when TMEM119 antibody was used as a huMG marker (Panel B, Supplementary Fig. 1a). Importantly, virtually all (>99.9%) of P2Y<sub>12</sub><sup>+</sup> cells expressed TMEM119, and >99.4% of TMEM119<sup>+</sup> cells expressed P2Y<sub>12</sub> in the FACS analysis (Supplementary Fig. 1b,c). To extend the phenotypic comparison between PBMCs, CSF cells and huMG, we manually gated CSF and blood mononuclear cells (i.e. monocytes, CD19<sup>+</sup> B cells, dendritic cells, CD56<sup>dim</sup><sup>+</sup> NK cells, CD3<sup>+</sup> T cells) and huMG (P2Y<sub>12</sub><sup>+</sup> or TMEM119<sup>+</sup>) clusters on the t-SNE map. On the basis of the mean signal intensity of all 57 markers analyzed (Panel A plus Panel B), huMG clustered distinctly from all cell subsets in blood and CSF (Fig. 1c). Specifically, huMG expressed higher mean levels of P2Y<sub>12</sub>, TMEM119, EMR1 (F4/80), CD64 and TREM2, whereas expression levels of CD44, CCR2, CD45, CD14 and CD16 were much lower in huMG compared with the PBMCs and CSF cells (Fig. 1d). In one huMG sample (out of 36 samples), we could detect an expression of CD19 or CD135 (Fig. 1c). However, these cells were positive for markers used to identify microglia such as P2Y<sub>12</sub>, TMEM119, EMR1, CX3CR1, CD11c and CD115. But they were negative for all classical B cell markers such as CCR7, CD62L, CD37 and CD40, as well as CD44, a general marker for peripheral immune cells. Therefore, a contamination of B cells or peripheral immune cells in this particular sample is improbable. Nevertheless, this rare expression of CD19 and CD135 deserves further investigation, which requires additional markers. CSF cells expressed comparatively high levels of the cytokines IL-6 and TNF- $\alpha$  (Supplementary Figs. 2 and 3). Interestingly, in all cell subsets, peripheral blood cells clustered separately from CSF cells (Fig. 1c). In sum, the unsupervised, t-SNE-based dimensionality reduction effectively demonstrated phenotypic segregation of huMG from circulating immune cells with or without the use of classical cell lineage markers, such as CD45, CD11b, CD3, CD19 and CD56 (Panel B, Supplementary Fig. 1). These tools allow large-scale collection and simultaneous immune profiling of huMG and circulating immune cells.

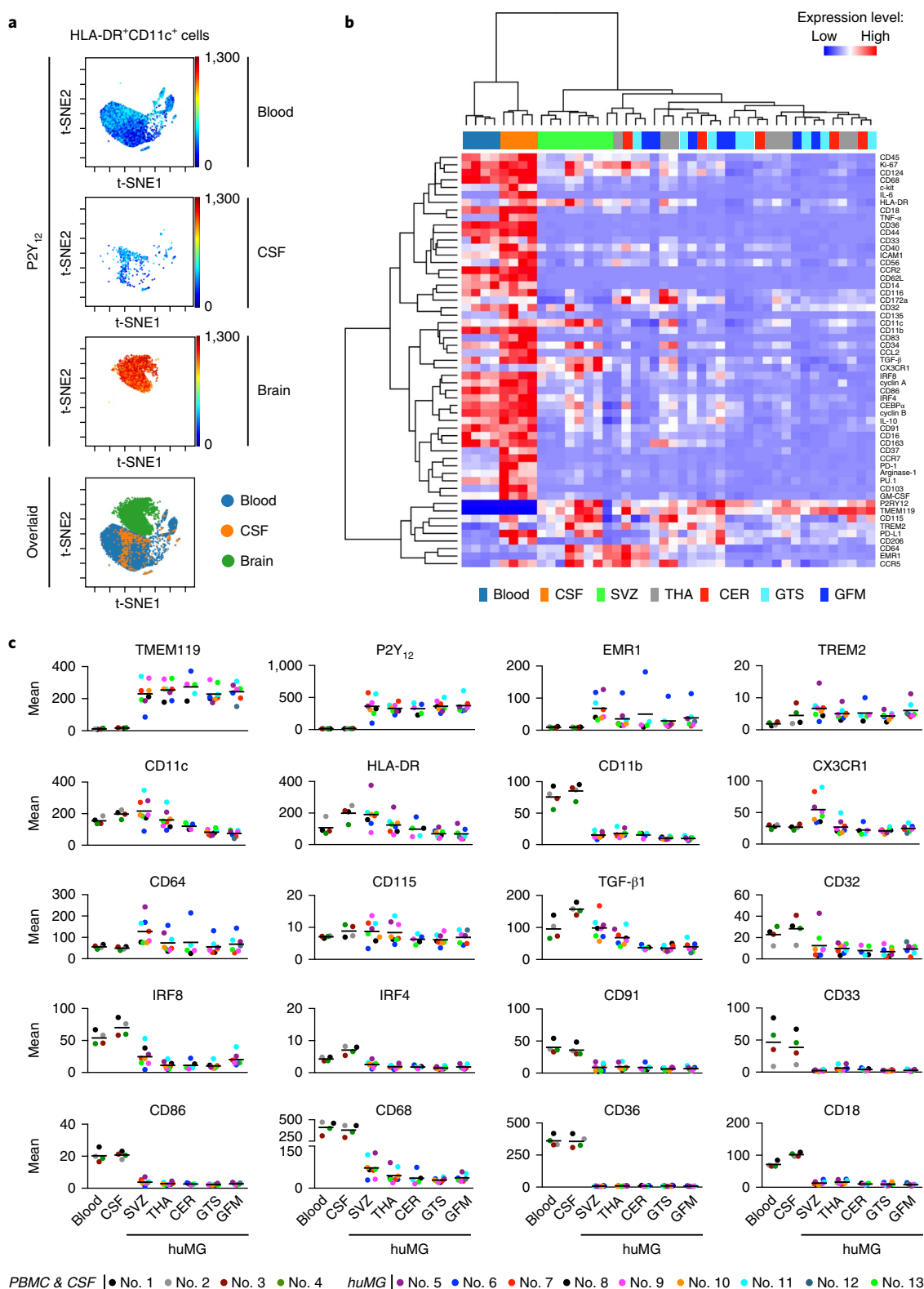
**Differential immunophenotypes of circulating myeloid cells and huMG.** Murine tissue-resident macrophages, including microglia, and circulating monocytes have distinct transcriptomic and enhancer landscapes that are regulated by the local microenvironment<sup>26</sup>. At the protein expression level, murine microglia can be separated from circulating monocytes and other tissue-resident macrophages by clustering<sup>27</sup>. Here, we observed that all P2Y<sub>12</sub><sup>-</sup> or TMEM119-expressing huMG also co-expressed HLA-DR and CD11c (Supplementary Fig. 4a,b). We therefore performed a

comprehensive t-SNE analysis of HLA-DR<sup>+</sup>CD11c<sup>+</sup> huMG and circulating myeloid cells (HLA-DR<sup>+</sup>CD11c<sup>+</sup>) from blood and CSF. The results shown in Fig. 2a indicate that huMG have a phenotype that distinguishes them from circulating myeloid cells in blood and CSF. We further analyzed these clusters at the level of marker expression on a donor-by-donor, compartment-by-compartment and brain region-by-region basis (Fig. 2b,c and Supplementary Figs. 2 and 3). On the basis of the expression levels of 55 investigated markers (Panel A plus Panel B, excluding CD3 and CD19), the heat map visualization in Fig. 2b revealed a distinct phenotypic signature of huMG, as well as significant phenotypic differences between the myeloid cell populations in blood and CSF from the same donors. P2Y<sub>12</sub> and TMEM119 expression was highly enriched in huMG and was absent from myeloid cells in blood and CSF (Fig. 2b,c).

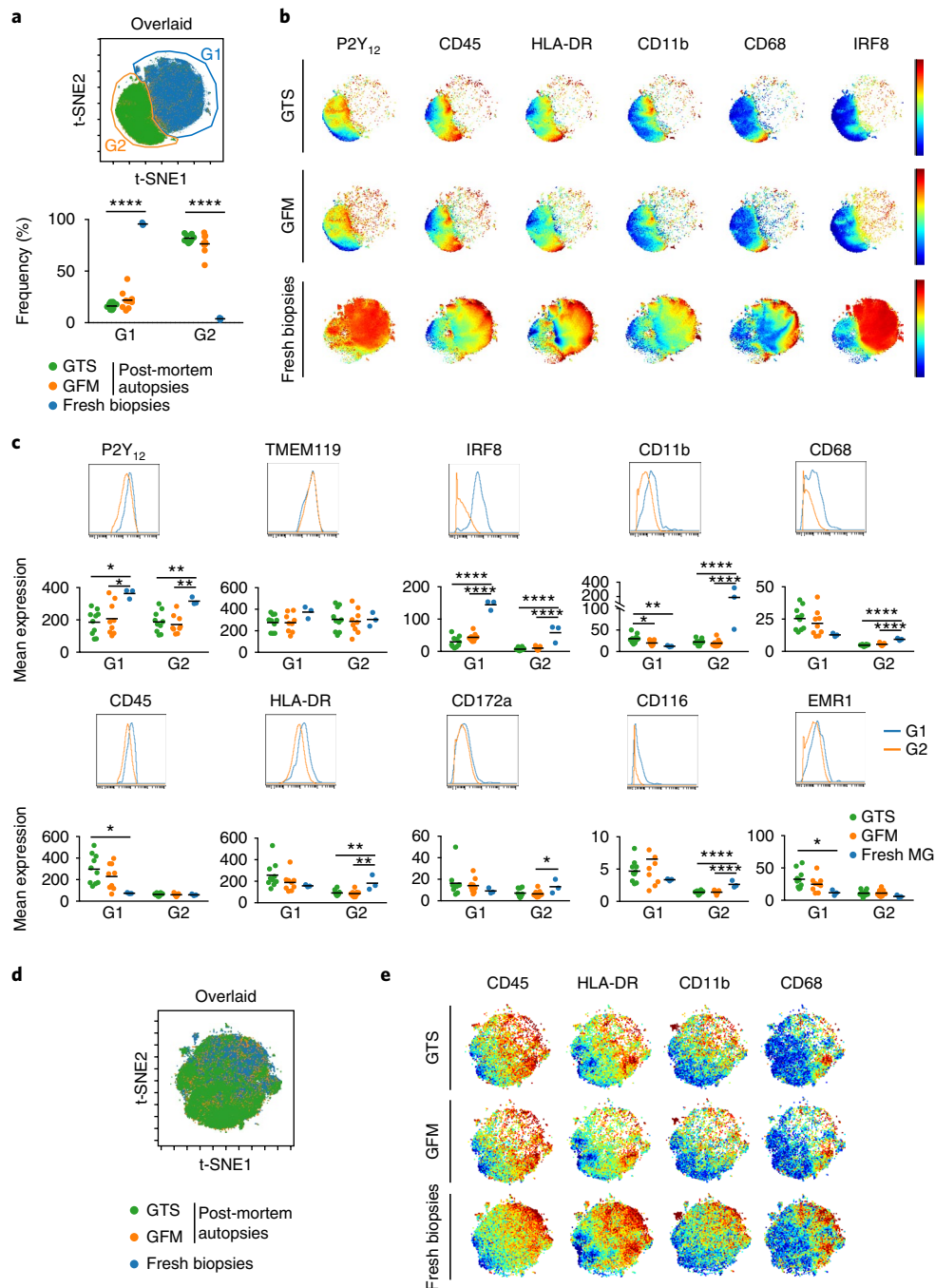
Next, we compared the immunophenotypic signatures of postmortem huMG with microglia isolated from temporal lobe biopsies (fresh huMG) (Supplementary Table 4). The biopsies were obtained from three patients during the resection of brain tissue for the treatment of epilepsy, analogous to the procedure recently described for transcriptomic and epigenetic profiling of huMG<sup>16</sup>. The epileptic focus with the strongest epileptogenic activity was removed before the surrounding tissue was used for microglia isolation. On the t-SNE map, postmortem huMG clustered together with fresh huMG and displayed a comparable immunophenotypic signature (Fig. 3a), underscoring the validity of using postmortem huMG for immunophenotypic profiling in health and disease. However, we did observe differences in the levels of signal intensity, in particular for IRF8 and P2Y<sub>12</sub>, and to a lesser degree for CD11b, CD68 and HLA-DR as well (Fig. 3b,c and Supplementary Fig. 5a,c). The expression levels of TMEM119 were not different between postmortem huMG and fresh huMG (Fig. 3c). Interestingly, we detected IRF8<sup>hi</sup>P2Y<sub>12</sub><sup>+</sup> cells (G1, Fig. 3a,b) at higher frequencies in fresh huMG than in postmortem huMG (Fig. 3c). Removing IRF8 and P2Y<sub>12</sub> from embedding parameters resulted in decreased phenotypic differences between postmortem huMG and fresh huMG (Fig. 3d,e and Supplementary Fig. 5b), suggesting that IRF8 and P2Y<sub>12</sub> are key markers that determine the difference between the two sources of huMG. Of note, these differences may result from postmortem changes and/or effects of epilepsy on the tissue.

#### Expression of mannose receptor C-type 1 (CD206) in huMG.

We observed variation in the expression of mannose receptor C-type 1 (MRC1 or CD206) in P2Y<sub>12</sub><sup>+</sup> cells across different brain regions (Fig. 4a,b). CD206 was previously suggested as a marker for M2 macrophages and perivascular macrophages in the human CNS<sup>28</sup>. However, recent data obtained from bulk RNA sequencing showed that huMG also express low levels of CD206 mRNA<sup>17,18</sup>. Moreover, activated murine microglia express CD206 after spinal cord injury<sup>29</sup>. In humans, it is unclear whether the expression of CD206 is confined to a subpopulation of microglia, and what the expression levels of CD206 are at the single-cell level. The low-dimensional t-SNE map of P2Y<sub>12</sub><sup>+</sup> cells derived from up to five brain regions from nine different donors showed a small cluster of CD206<sup>high</sup>CD163<sup>+</sup>CD14<sup>+</sup> cells (G3), which highly expressed HLA-DR, CD68 and CD11b (Fig. 4c,d), suggestive of perivascular macrophages (pm $\Phi$ ). Interestingly, we observed low expression of P2Y<sub>12</sub> on this population (Fig. 4c,d and Supplementary Fig. 6a), which has not been reported before<sup>12,20</sup>. However, the expression levels of P2Y<sub>12</sub> were much lower than for huMG (Supplementary Fig. 6a). The cluster of pm $\Phi$  could not be detected when antibodies against CD163, CD14, CD68 and CD11b were not included in the staining panel (Panel B; Supplementary Fig. 7a). Interestingly, we also detected other clusters of CD206<sup>low</sup>CD163<sup>-</sup>CD14<sup>-</sup>P2Y<sub>12</sub><sup>high</sup> cells (G2 and G3; Fig. 4c,d). The CD206<sup>low</sup> huMG did not express CD163, and showed lower expression levels of HLA-DR, CD68, CD33, CD11b and CD45 compared with pm $\Phi$  (Fig. 4c,d).



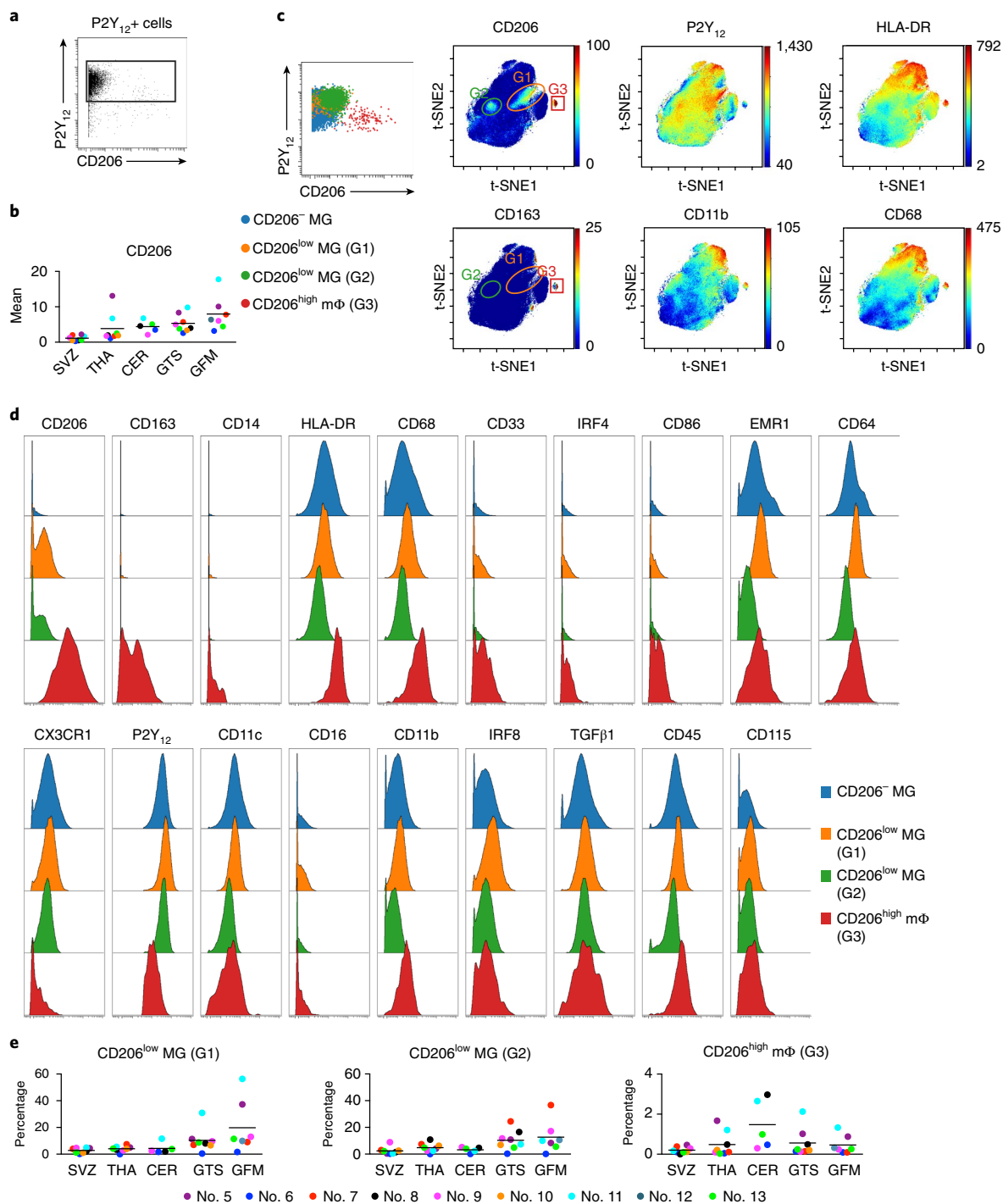
**Fig. 2 | Comparative phenotypic analysis of huMG and peripheral myeloid cells.** **a**, Representative t-SNE projection of multidimensional single-cell phenotypes of HLA-DR<sup>+</sup>CD11c<sup>+</sup> myeloid cells detected in blood (PBMCs, *n* = 4 biologically independent samples), CSF (*n* = 4 biologically independent samples) and brain (huMG, *n* = 36 biologically independent samples). The color spectrum represents an expression level of P2Y<sub>12</sub> (red, high expression; dark blue, no expression). The bottom image shows an overlaid t-SNE plot of all cells from all three compartments (blue, blood; orange, CSF; and green, brain huMG). **b**, Heat map cluster demonstrates the mean expression of all 55 markers (Panel A plus Panel B) and relationships between blood (blue), CSF (orange) and brain myeloid cells (SVZ, bright green; THA, gray; CER, red; GTS, bright blue; GFM, dark blue). Heat colors have been scaled per marker (red, high expression; blue, low expression). **c**, Mean expression levels of selected markers in blood, CSF and brain myeloid cells (black lines show mean values of the datasets).



**Fig. 3 | Comparative immune profiling of postmortem huMG and fresh huMG. a**, An overlaid t-SNE plot of all cells from all samples (green, postmortem GTS-huMG,  $n = 10$  biologically independent samples; orange, postmortem GFM-huMG,  $n = 9$  biologically independent samples; blue, huMG from fresh biopsies,  $n = 3$  biologically independent samples). Two main clusters, G1 (blue) and G2 (orange), are detected. The graph below shows the quantitative frequencies of G1 and G2 populations in each sample (black lines show mean values of the datasets). \*\*\*\* $P < 0.0001$ , one-way ANOVA with Bonferroni correction. **b**, Representative reduced-dimensional single-cell t-SNE illustrations of P2Y<sub>12</sub><sup>+</sup> huMG from biologically independent samples of GTS ( $n = 10$ ), GFM ( $n = 9$ ) and fresh biopsies ( $n = 3$ ). The color spectrum represents an expression level (red, high expression; dark blue, low expression). **c**, Mean expression levels of selected markers showing differential marker expressions between the two gates (G1 and G2 in **a**) in huMG from biologically independent samples of GTS (green,  $n = 10$ ), GFM (orange,  $n = 9$ ) and fresh biopsies (blue,  $n = 3$ ). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , one-way ANOVA with Bonferroni correction. **d**, An overlaid high-dimensional plot (embedding without P2Y<sub>12</sub> and IRF8) of all cells from all biologically independent samples (green, GTS,  $n = 10$ ; orange, GFM,  $n = 9$ ; blue, fresh biopsies,  $n = 3$ ). No distinct cluster was detected, thus differences between samples are minute. **e**, Representative reduced-dimensional single-cell t-SNE illustration of P2Y<sub>12</sub><sup>+</sup> huMG from biologically independent samples of GTS ( $n = 10$ ), GFM ( $n = 9$ ) and fresh biopsies ( $n = 3$ ). The color spectrum represents an expression level of CD45, HLA-DR, CD11b and CD68 (red, high expression; dark blue, low expression).

However, CD206<sup>low</sup> huMG showed a higher expression level of CX3CR1 than pm $\Phi$  (Fig. 4d). In Panel B, the clusters of CD206<sup>low</sup> huMG were also identified (Supplementary Fig. 7a,b).

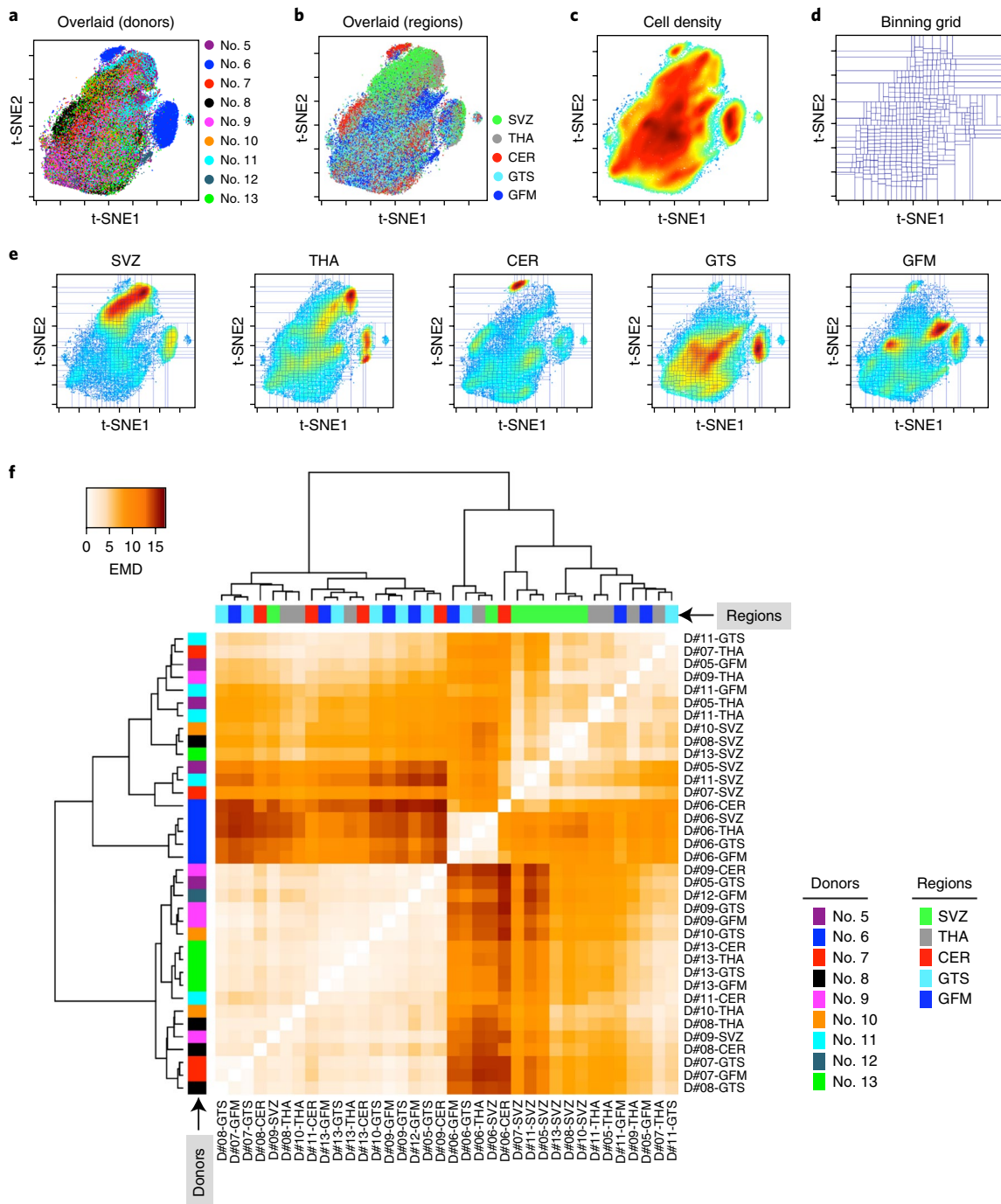
Interestingly, we observed regional heterogeneity in the distribution of CD206<sup>low</sup> huMG. Quantification of the manually gated populations G1–G3 on the t-SNE display revealed that CD206<sup>low</sup>



**Fig. 4 | CD206-expressing huMG and perivascular macrophages in the CNS mononuclear cell fraction.** **a**, Representative two-dimensional dot plot of brain mononuclear cell fraction (two independently repeated experiments with similar results;  $n = 36$  biologically independent samples) showing cell population co-expressing P2Y<sub>12</sub> and CD206. **b**, Mean expression level of CD206 in P2Y<sub>12</sub><sup>+</sup> cells gated in Fig. 1a (biologically independent samples of SVZ,  $n = 8$ ; THA,  $n = 8$ ; CER,  $n = 5$ ; GTS,  $n = 8$  and GFM,  $n = 7$ ). Black lines show the mean of datasets. **c**, High-dimensional t-SNE plots of concatenated FCS file (all 36 huMG samples). Each dot represents one cell. The color spectrum represents an expression level of CD206, P2Y<sub>12</sub>, HLA-DR, CD163, CD11b and CD68 (red, high expression; blue, no expression). CD206<sup>high</sup> perivascular mΦ is gated as ‘G3’ (red square), and CD206<sup>low</sup> cell population is gated as ‘G1’ (orange circle) and ‘G2’ (green circle). **d**, Histogram plots show an expression of selected markers in CD206<sup>-</sup> (blue), CD206<sup>low</sup> (G1, orange; G2, green) huMG and CD206<sup>high</sup> perivascular mΦ (G3, red). **e**, Frequencies of each CD206-expressing population in different brain regions. The values of an individual donor were plotted in the same color. The black lines represent the mean value.

huMG were more frequent in temporal lobe (gyrus temporalis superior, GTS) and frontal lobe (gyrus frontalis medius, GFM) compared with other brain regions, whereas CD206<sup>high</sup> pmΦ were

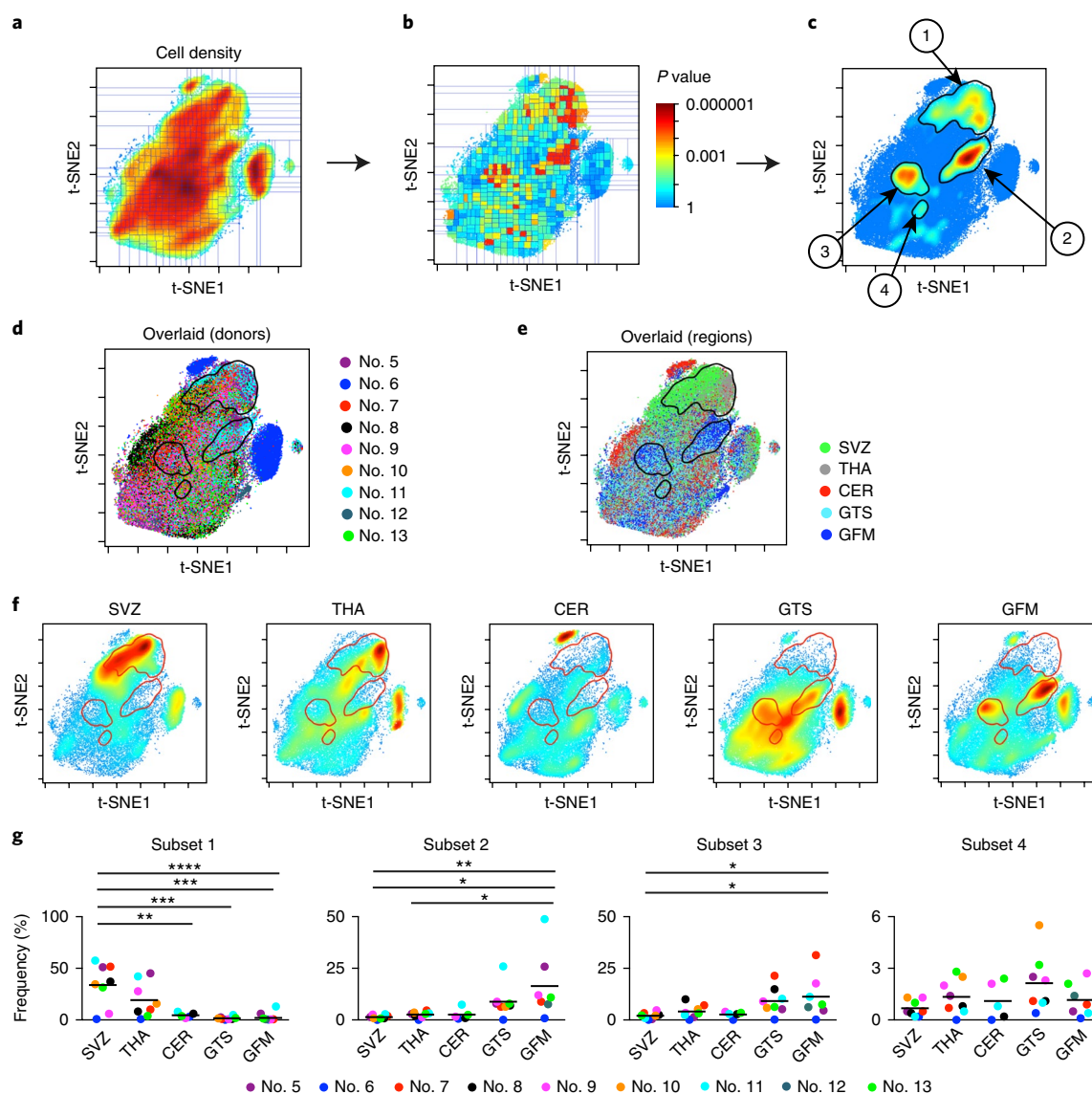
equally distributed across the human brain (Fig. 4e). High proportions of CD206<sup>low</sup> huMG in temporal and frontal lobes were confirmed using Panel B (Supplementary Fig. 7b). Of note, such



**Fig. 5 | Assessment of regional differences in huMG phenotypes by probability binning.** The top panel shows the same t-SNE plot of concatenated FCS files from 36 huMG samples. **a–c**, The coloring indicates nine donors (**a**), five brain regions (**b**), or overall cellular density of the concatenated files (**c**) (spectrum from blue (low density) to red (high density)). **d**, From the concatenated files a single binning grid is established, comprising 512 ‘microgates’ in which cell frequencies are enumerated. Superimposing the binning grid on the t-SNE landscapes of individual samples allows comparative analysis of huMG profiles at a sample-to-sample basis across all five regions. **e**, t-SNE plots of concatenated FCS files of each brain region. **f**, Heat map representing the pairwise EMD between cellular density distributions over the t-SNE space among all huMG samples. Hierarchical clustering highlights samples that have a highly similar phenotype, as indicated by low EMD values. The values of EMD range from 0 to 17, as shown by the color bar (top left).

low expression of CD206 on P2Y<sub>12</sub>-expressing cells failed to be detected by flow cytometry due to the high autofluorescent background of postmortem huMG (Supplementary Fig. 6b). The results underscore the power of the multidimensionality of mass cytometry in attempts to identify microglia subpopulations in the human brain.

**Heterogeneity of human postmortem brain microglia.** Recently, region-dependent microglial diversity was detected in the mouse brain on the basis of transcriptional profiling using microarrays<sup>14</sup>. Here, we studied the phenotypic signatures of huMG at the single-cell level, and addressed the issue of regional heterogeneity of human brain microglia by mass cytometry. Our initial results



**Fig. 6 | Region-dependent huMG subpopulations.** **a**, Binning grid superimposed on concatenated t-SNE map ( $n = 36$  biologically independent samples). **b**, Statistical t-SNE map resulting from bin-wise testing for frequency differences between huMG samples ( $n = 35$  biologically independent samples) from five brain regions (SVZ,  $n = 8$ ; THA,  $n = 8$ ; CER,  $n = 5$ ; GTS,  $n = 8$ ; GFM,  $n = 6$ ) using the non-parametric Skillings–Mack statistic for unbalanced two-way block designs. The color spectrum corresponds to unadjusted  $P$  values, ranging from  $<0.000001$  (red) to 1 (blue). **c**, Smoothed representation of statistical t-SNE map in **b** after FDR adjustment for multiple comparisons and thresholding to 0.05 FDR-adjusted  $P$  values (blue). Areas throughout the t-SNE map comprising significant bins (green to red spectrum) are indicative of differentially abundant subsets with distinct phenotypes and detected using a density-based automated gating approach (black contour lines). **d–f**, Gates of four detected subsets (black lines) are shown in concatenated t-SNE maps, color encoded to indicate nine donors (**d**), five brain regions (**e**) and in concatenated cellular density plots for each of the five investigated brain regions (**f**), individually. **g**, The graphs show frequencies (%) of all four differentially abundant subsets across five brain regions (biologically independent samples of SVZ,  $n = 8$ ; THA,  $n = 8$ ; CER,  $n = 5$ ; GTS,  $n = 8$ ; GFM,  $n = 7$ ). Black lines show mean of the dataset.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ , one-way ANOVA with Bonferroni correction.

suggested that huMG from the subventricular zone (SVZ) display a phenotype that is distinct from huMG in other brain regions (Figs 1c and 2b). To extend these observations, we performed a comprehensive cluster analysis using t-SNE embedding of the entire data set, including all brain regions and all donors (36 samples; Fig. 5a–c). To quantify phenotypic differences and to fully harness the multi-dimensional nature of the mass cytometry data, we combined the t-SNE algorithm with probability binning<sup>30</sup>. The binning model was created on collapsed data from all samples (that is a concatenated flow cytometry standard (FCS) file) by recursively splitting the events at the median values along the two t-SNE dimensions to

yield 512 microgates (binning grids) at sufficiently high resolution (Fig. 5d,e). We used the earth mover's distance (EMD) metric<sup>31</sup> to quantify cell-distributional differences between huMG of samples from different donors and brain regions (Fig. 5f). The EMD score between most of the huMG in the SVZ (six of eight investigated donors) was very low, suggesting strong similarity between the SVZ samples (brain region) rather than donor-specific huMG phenotype (see also Supplementary Table 5). Using this methodology, huMG in the SVZ were confirmed to be phenotypically distinct from huMG in other brain regions (Fig. 5f). However, we also observed donor-dependent phenotypic variability (Fig. 5a and Supplementary



Fig. 8a–c). In particular, donor number 6 revealed a distinct cluster of huMG that we further characterized as a CD64<sup>hi</sup>EMR1<sup>hi</sup> population (Supplementary Fig. 8b,c). Removing this donor or the outlier markers (CD64 and EMR1) before t-SNE embedding did not change our results regarding the regional heterogeneity of huMG (Supplementary Fig. 8d,e). We therefore included the outlier donor (number 6) and the outlier markers (CD64 and EMR1) in all further analyses so as to embrace the biological variability of huMG.

To determine and visualize frequencies of differential phenotypes between brain regions, we performed bin-wise, intrasubject, mass univariate statistical testing<sup>32</sup> using the Skillings–Mack Friedman-type nonparametric one-way repeated measures statistic<sup>33</sup> to account for the non-normality of cell frequency data, the incomplete block design (namely, unequal number of brain regions between investigated donors) and the small sample sizes. The results of the group-level analysis are presented as a single statistical t-SNE map (Fig. 6a,b and Supplementary Fig. 8f,g), in which areas of connected bins exceeding a given significance threshold are automatically gated to reveal cellular phenotypes accounting for the detected differences. Using this analysis, we identified four huMG subsets that showed differential abundance in different brain regions (subsets 1, 2 and 3:  $P < 0.0001$ ; subset 4:  $P = 0.0014$ , Skillings–Mack test with controlled false discovery rate (FDR)) (Fig. 6c–e). We observed that the SVZ and thalamus (THA) contain similar huMG phenotypes (subset 1), which are virtually absent from the other brain regions (Fig. 6e–g and Supplementary Fig. 8f,g). Temporal lobe and frontal lobe are enriched in different huMG phenotypes (subsets 2–4). Interestingly, subset 4 appears to be more abundant in the temporal lobe than in the frontal lobe (Fig. 6e–g). The profile of huMG in the cerebellum (CER) was distinct from the other brain regions and revealed low abundance of all subsets (Fig. 6e–g). Similar regional differences were also detected when antibody Panel B was applied to the samples (Supplementary Figs. 9 and 10). Finally, we confirmed our findings using the differential abundance hypersphere analysis in original multiparameter space with the cydar/edgeR framework<sup>34</sup> (Supplementary Figs. 8h and 9c).

**Region-dependent phenotypic signatures of huMG.** Next, we further characterized the phenotypic signatures of the huMG regional subsets identified in Fig. 6. The four subsets were automatically gated and profiled for marker expression. The phenotypic signature of each subset was extracted (Fig. 7a and Supplementary Fig. 10). Subset 1, which was observed in higher proportions in the SVZ and THA, showed higher expression of CD11c, CD195 (CCR5), CD45, CD64, CD68, CX3CR1, EMR1 and HLA-DR compared with the other subsets (Fig. 7a,b and Supplementary Fig. 10b,d). Moreover,

subset 1 of huMG expressed higher levels of the proliferation markers cyclin A, cyclin B1 and Ki67 (Supplementary Fig. 8i). These features suggest a more activated state of microglial cells in the SVZ and thalamus. Subsets 2 and 3, which were more abundant in GTS and GFM, expressed higher levels of CD206 compared with the other subsets (Fig. 7a,b and Supplementary Fig. 10b,d). Although the two subsets were generally very similar, subsets 2 and 3 differed in their expression of CD64 and EMR1 (Fig. 7a,b). Interestingly, we observed a positive correlation between donor age and the expression of CD11b, CD68, CD64, HLA-DR and TREM2 in huMG from different brain regions, although the results need to be interpreted with caution given the small sample size (Supplementary Fig. 11). Importantly, we identified CD11c, CD206, CD45, CD64, CD68, CX3CR1, HLA-DR and IRF8 as key markers for the detection of huMG regional heterogeneity (Fig. 7b). To test the feasibility of using these eight molecules in a reduced binary panel and conventional gating, we applied sequential (that is, Boolean) gating strategies to identify the four putative subsets on the basis of the expression of only these eight markers (Fig. 7c). Then, we compared the outcome frequencies of each Boolean-gated subset with the frequencies obtained by the gates in the t-SNE plot shown in Fig. 6. The frequencies of subsets 1, 2 and 3 were comparable between the two approaches (subset 1:  $\rho/r^2 = 0.87/0.91$ ; subset 2:  $\rho/r^2 = 0.85/0.78$ ; subset 3:  $\rho/r^2 = 0.87/0.77$ , Fig. 7d), whereas the frequencies of the lower-abundant subset 4 were slightly different between the two types of analysis (subset 4:  $\rho/r^2 = 0.72/0.34$ , Fig. 7d). We confirmed the suitability of these eight markers to identify huMG subsets 1, 2 and 3, whereas the detection of the rare subset 4 remained challenging. Next, we tested the robustness of the eight defined markers for identifying regional huMG heterogeneity using the flowType/RchyOptimyx pipeline<sup>35</sup> (Fig. 7e). Populations with highest  $-\log_2(P)$  significance scores highlight CD206, CD45, CD64, CD68 and HLA-DR as the most important markers to target region-specific huMG phenotypes in a manual gating strategy.

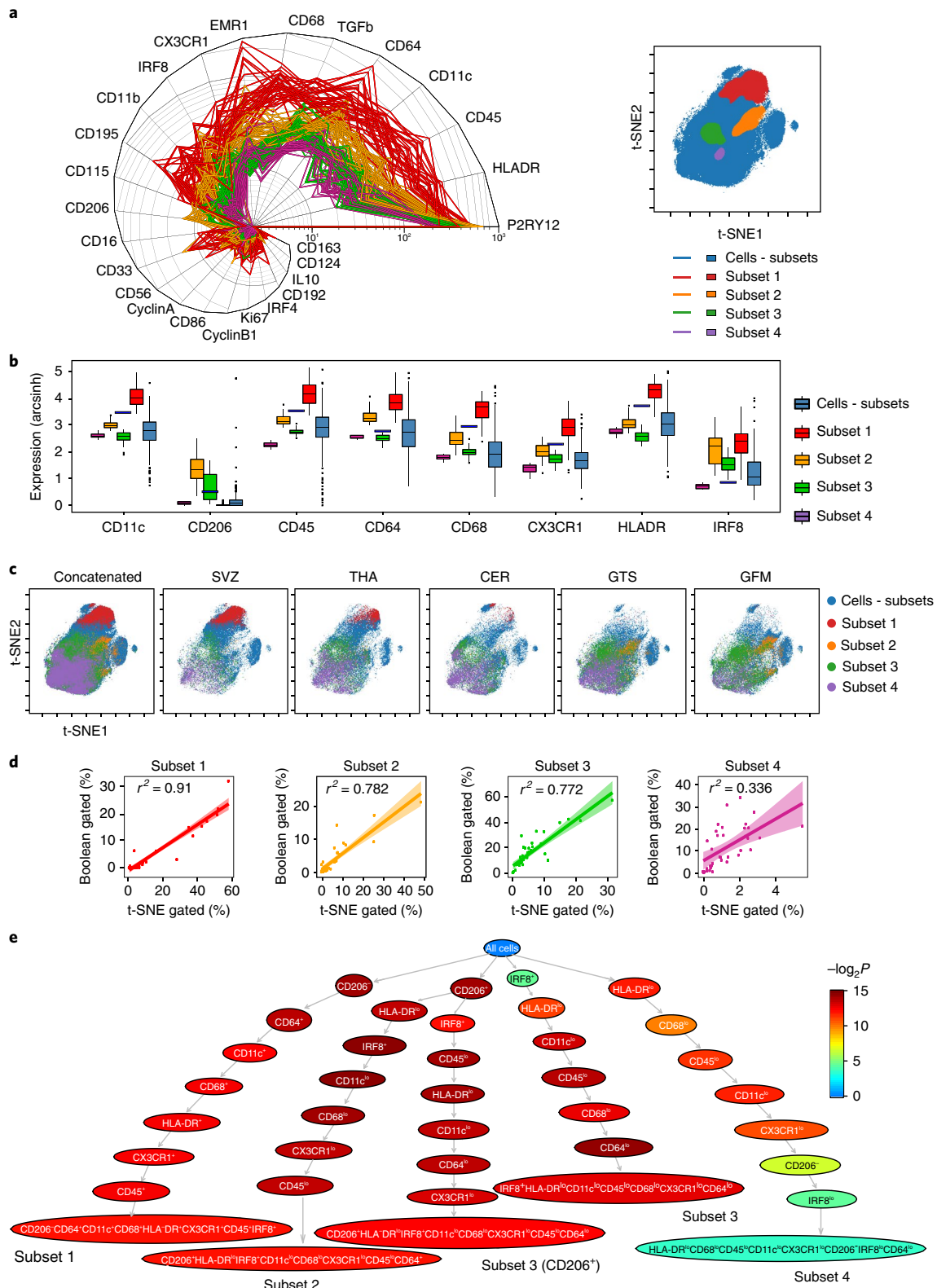
Finally, we performed FACS analysis to test the feasibility of identifying regional heterogeneity of huMG (Supplementary Table 4) by a more widely available technology. We were able to detect phenotypic differences of huMG between regions (SVZ versus GTS and GFM) on the basis of the FACS analysis of CD45, CD64, CD68 and HLA-DR expression (Fig. 8a). Furthermore, we detected a cluster analogous to CyTOF subset 1 (Figs. 6 and 7) of huMG (Fig. 8b, red gate). Importantly, the frequency of this SVZ-enriched subset was comparable between FACS analysis using four markers (CD45, CD64, CD68 and HLA-DR) and the CyTOF measurement (Fig. 8c).

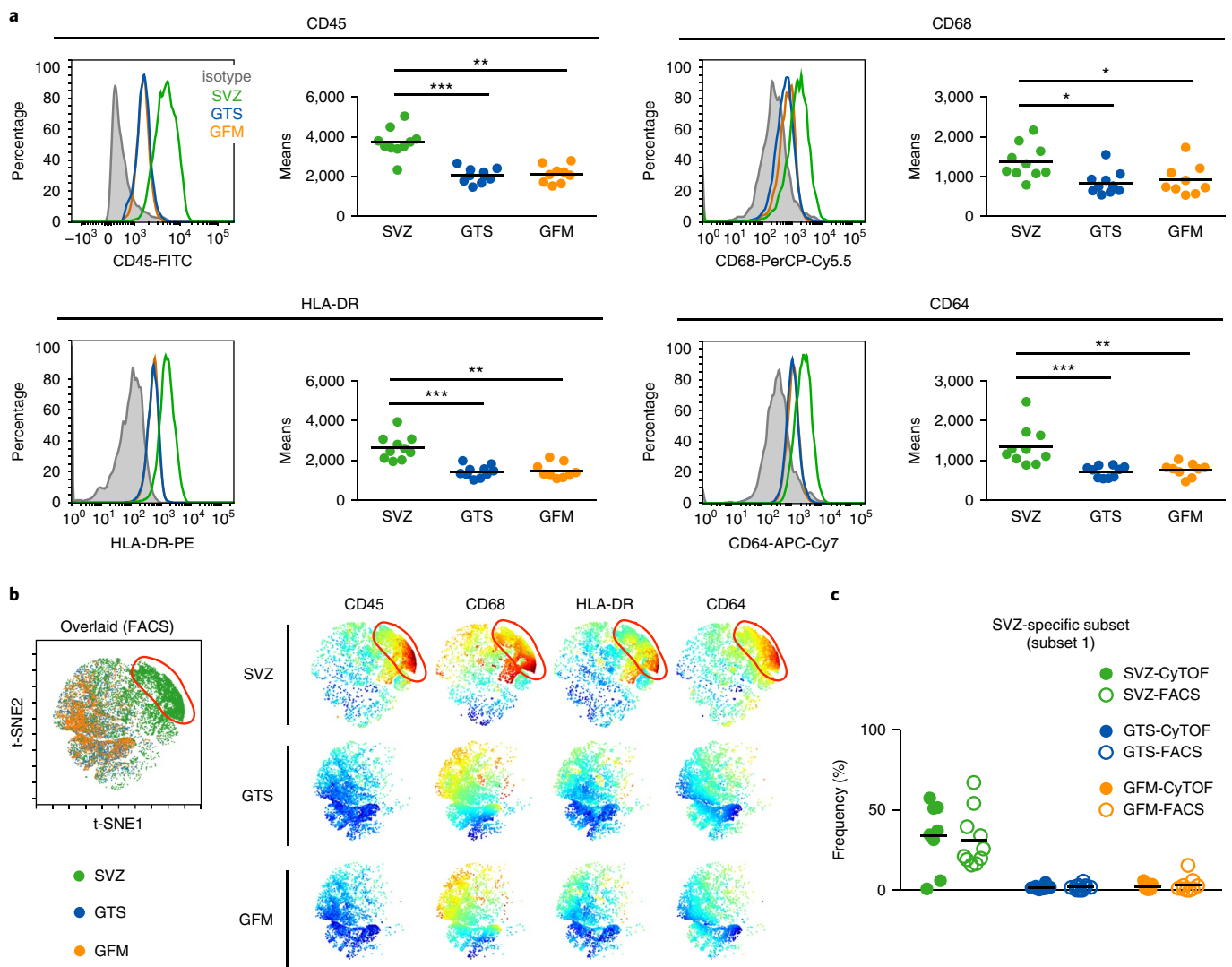
**Fig. 7 | Region-dependent huMG phenotypes.** **a**, Radar chart (or snail plot) shows marker expression levels of each huMG subset (subset 1, red; subset 2, orange; subset 3, green; subset 4, purple). The snail shell represents transverse (perpendicular) axis mapping marker expression levels on an exponential scale. Each line denotes each sample ( $n = 36$  biologically independent samples). The right image demonstrates the automated subset gating on the t-SNE map (subset 1, red; subset 2, orange; subset 3, green; subset 4, purple; remaining cells (cells – subsets), all blue dots). **b**, Selection of eight markers defining huMG subsets ( $n = 36$  biologically independent samples). Median bin expression levels are shown for each subset and marker in box plot representation (subset 1, red box,  $n = 66$  bins; subset 2, orange box,  $n = 32$  bins; subset 3, green box,  $n = 33$  bins; subset 4, purple box,  $n = 7$  bins; remaining cells (cells – subsets), blue box,  $n = 374$  bins). Blue lines indicate phenotype-defining cutoffs, used to identify the subsets by conventional (multivariate or hierarchical) gating. Box center and limits represent median, upper and lower quartiles; whiskers define the  $1.5 \times$  interquartile range; points show outliers. **c**, t-SNE plots of concatenated FCS files (from left to right: biologically independent samples of all,  $n = 36$ ; SVZ,  $n = 8$ ; THA,  $n = 8$ ; CER,  $n = 5$ ; GTS,  $n = 8$ ; GFM,  $n = 7$ ) are overlaid with the four subsets identified by multivariate gating using a (Boolean) combination of one-dimensional gates set according to cutoffs and markers shown in **b**. **d**, Scatter plots showing correlation between subset frequencies detected by Boolean gating and by automated t-SNE gating in  $n = 36$  biologically independent samples. Shaded areas indicate 0.95 confidence intervals of the linear regression,  $r^2$  denotes respective coefficients of determination. **e**, Marker importance analysis using the flowType/RchyOptimyx pipeline. The graph shows optimized gating hierarchies of the subsets starting from ungated cells (top node) to the eight-marker phenotypes (bottom nodes) as defined by cutoffs in **b**. The color of the nodes shows significance scores of brain region-dependent differential abundance as the negated  $\log P$  value of the Skillings–Mack test conducted on the same  $n = 35$  independent huMG samples (SVZ,  $n = 8$ ; THA,  $n = 8$ ; CER,  $n = 5$ ; GTS,  $n = 8$ ; GFM,  $n = 6$ ) from eight individual donors for each preferential addition of a subset-defining marker (node labels) which contributes at most to an increase in the significance score. The CD206<sup>+</sup> subset 3 phenotype has been included to also target the fraction of cells with CD206 expression above cutoff.

Discussion

Microglia are resident innate immune cells in the human CNS that are involved in neural development and function, as well as responses to diseases. Although rodent microglia are often used to investigate microglial function, the emerging differences between human and rodent microglia call into question the clinical relevance of some

of the research findings obtained in laboratory animals<sup>36,37</sup>. Several research groups have established a firm basis for the use of huMG in neuroscience and neuroimmunology and have provided invaluable transcriptomic information on these cells<sup>12,18,19,38–44</sup>. However, phenotypic profiling of huMG on the basis of a comprehensive array of marker proteins has remained technically challenging.





**Fig. 8 | Identification of huMG heterogeneity by flow cytometry.** **a**, HuMG isolated from SVZ ( $n=10$ ; green dots/lines) express higher levels of CD45, CD68, HLA-DR and CD64 than the ones isolated from GTS ( $n=10$ ; blue dots/lines) and GFM ( $n=9$ ; orange dots/lines). Data show mean, independent samples, each from an individual donor (total of ten donors). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , one-way ANOVA with Bonferroni correction. **b**, Representative multidimensional single-cell t-SNE projections of huMG from biologically independent samples of SVZ ( $n=10$ ), GTS ( $n=10$ ) and GFM ( $n=9$ ). The color spectrum represents an expression levels of CD45, CD68, HLA-DR and CD64 (red, high expression; dark blue, low expression). The left image shows overlaid multidimensional plot of all cells from all three brain regions (green, SVZ; blue, GTS; orange, GFM). **c**, Quantification of the frequency of defined SVZ-specific subset (G1 and red gate in **b**) obtained by FACS compared with the frequency of subset 1 quantified by CyTOF (as shown in Fig. 6g). CyTOF: SVZ ( $n=8$ ), GTS ( $n=8$ ) and GFM ( $n=7$ ); FACS: SVZ ( $n=10$ ), GTS ( $n=10$ ) and GFM ( $n=9$ ). No significant differences were found between the two techniques (multiple unpaired t-test, two-sided, corrected for multiple comparison using the Sidak-Bonferroni method).  $P=0.7906$  (SVZ);  $0.7471$  (GTS) and  $0.4767$  (GFM). Black lines show the mean of datasets.

This is particularly true at the single-cell level, which is required to identify microglia subpopulations.

Here, we have applied massive single-cell immune profiling of huMG from different brain regions by multiplexed mass cytometry, allowing for a detailed phenotypic characterization of huMG. Our findings support the notion of microglial heterogeneity in the human brain, which is in line with recent data obtained from mouse microglia<sup>14</sup>. Our results substantiate previously published data on messenger RNA<sup>12,16–18</sup> and protein<sup>12,18–20,44,45</sup> expression of huMG. The development of a new cell fixation and cryopreservation technique combined with barcoding (multiplexing), mass cytometry and novel algorithms for data analysis enabled us to identify a phenotypic signature of huMG that distinguishes them from other innate immune cells (for example, cells from blood and CSF).

In this study, we combined t-SNE and probability binning for both of their strengths to detect changes in subsets that are defined by dim or unimodal marker (co-) expression or subtle shifts in expression levels (and thus difficult to enumerate by clustering techniques). This approach revealed microglial subsets that differ in their abundance across different regions of the human brain, indicating phenotypic heterogeneity among huMG. Notably, we extracted a panel of eight (out of 57) phenotypical markers that allow us to distinguish huMG subsets.

We confirmed the key transcriptomic signature of huMG<sup>16,17</sup> at the protein level, namely, the expression of P2Y<sub>12</sub> and TMEM119, the high expression of CD64, CX3CR1, TGF- $\beta$ 1, TREM2, CD115, CCR5, CD32, CD172a and CD91, and the low to absent expression of CD44, CCR2, CD45, CD206, CD163, and CD274 (PD-L1).

The results are in line with recent mass cytometry data on immune cells in the mouse brain, which reveal that CD44 is expressed only on infiltrating cells and not on resident myeloid cells<sup>46</sup>. These core immunophenotypes of postmortem huMG are apparently comparable with fresh huMG, albeit with differences in signal intensities for some markers. Szulzewsky et al.<sup>47</sup> observed similar transcriptomic profiles between huMG from epilepsy and postmortem tissues. Interestingly, we detected the expression of an EGF-like module containing mucin-like hormone receptor (EMR)1, the human ortholog of F4/80, in huMG, whereas circulating monocytes and myeloid dendritic cells in the blood and CSF lacked EMR1 expression. This is a surprising finding given that EMR1 has been suggested to be a highly specific marker for eosinophils in human and is absent on mononuclear phagocytic cells including monocytes, macrophages and dendritic cell subsets<sup>48</sup>.

The combination of antibody panel design, single-cell mass cytometry and computational unsupervised data analysis separated cells with characteristics of perivascular macrophages (CD11b<sup>+</sup>CD206<sup>high</sup>CD163<sup>+</sup>) from huMG (CD11b<sup>+</sup>CD206<sup>low</sup>–CD163<sup>–</sup>). Furthermore, we identified two microglial subsets that express CD206, but not CD163, particularly in the frontal and temporal lobe. These findings are in line with recent studies that showed low expression of CD206 mRNA in human microglia<sup>17,18</sup>. Notably, microglia were found to express CD206 after spinal cord injury in mice<sup>29</sup>, suggesting that CD206 expression may reflect the functional responses of huMG.

Recent studies demonstrate that microglia in mice and humans are short lived and quickly renewed at the individual cell level<sup>49,50</sup>. At the population level, CNS microglia are maintained by the balanced regulation between proliferation and apoptosis. In agreement with these observations, we were able to detect proliferating huMG at different phases of the cell cycle across five brain regions. Of note, we observed slightly higher expression of Ki-67 (G1, S & G2 phase, mitosis), cyclin A (S and G2 phase, mitosis) and cyclin B (mitosis) on huMG subsets in the SVZ and thalamus compared with other brain regions. Whether this increased proliferation mirrors region-specific phenotype and/or function of huMG remains to be investigated.

One important conclusion drawn from this study is that huMG have multiple phenotypic signatures that appear to depend on the brain region that they reside in, whereas the core immunophenotype that distinguishes them from circulating and/or infiltrating myeloid cells is retained postmortem across the five investigated regions of human brain. The results are in line with previous findings in mouse and human<sup>14,18,19</sup>. We detected comparatively higher expression of markers involved in microglial activation, such as CD68, CD86, CD45, CX3CR1, CD11c, CD64, ERM1 and HLA-DR in the SVZ and thalamus compared with other brain regions. Interestingly, huMG subpopulations in the temporal and frontal lobe expressed low levels of the mannose receptor CD206, which is a marker of M2-polarized macrophages. Whether this phenotype implies a region-specific function remains to be investigated. The extent to which the isolation protocol could influence the observed differences in regional expression profile is unclear, although the finding that huMG in the THA and SVZ have similar phenotypic profiles despite differences in the isolation protocol suggests this may not be a major factor.

In conclusion, this study demonstrates the power of combining multiplexed mass cytometry with bioinformatics to reveal region-dependent signatures of huMG, even for small sample size and/or when the differences between groups are very small. We believe that an appropriate protocol for sample preparation is one of the key success factors for immune profiling of human postmortem microglia. Our findings of microglial heterogeneity in the human brain may help to reveal region-specific functions of these cells in health and disease, and instruct the development of more selective pharmacological interventions targeting microglia in humans.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41593-018-0290-2>.

Received: 6 October 2017; Accepted: 13 November 2018;  
Published online: 17 December 2018

## References

- Prinz, M. & Priller, J. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat. Rev. Neurosci.* **15**, 300–312 (2014).
- Sierra, A. et al. Microglia shape adult hippocampal neurogenesis through apoptosis-coupled phagocytosis. *Cell Stem Cell* **7**, 483–495 (2010).
- Parkhurst, C. N. et al. Microglia promote learning-dependent synapse formation through brain-derived neurotrophic factor. *Cell* **155**, 1596–1609 (2013).
- Prinz, M. & Priller, J. The role of peripheral immune cells in the CNS in steady state and disease. *Nat. Neurosci.* **20**, 136–144 (2017).
- Perry, V. H. & Holmes, C. Microglial priming in neurodegenerative disease. *Nat. Rev. Neurol.* **10**, 217–224 (2014).
- Colonna, M. & Butovsky, O. Microglia function in the central nervous system during health and neurodegeneration. *Annu. Rev. Immunol.* **35**, 441–468 (2017).
- Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).
- Ginhoux, F. et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
- Kierdorf, K. et al. Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16**, 273–280 (2013).
- Elmore, M. R. et al. Colony-stimulating factor 1 receptor signaling is necessary for microglia viability, unmasking a microglia progenitor cell in the adult brain. *Neuron* **82**, 380–397 (2014).
- Bruttger, J. et al. Genetic cell ablation reveals clusters of local self-renewing microglia in the mammalian central nervous system. *Immunity* **43**, 92–106 (2015).
- Butovsky, O. et al. Identification of a unique TGF- $\beta$ -dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
- Orre, M. et al. Acute isolation and transcriptome characterization of cortical astrocytes and microglia from young and aged mice. *Neurobiol. Aging* **35**, 1–14 (2014).
- Grabert, K. et al. Microglial brain region-dependent diversity and selective regional sensitivities to aging. *Nat. Neurosci.* **19**, 504–516 (2016).
- Orre, M. et al. Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. *Neurobiol. Aging* **35**, 2746–2760 (2014).
- Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).
- Galatro, T. F. et al. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat. Neurosci.* **20**, 1162–1171 (2017).
- Melief, J. et al. Characterizing primary human microglia: a comparative study with myeloid subsets and culture models. *Glia* **64**, 1857–1868 (2016).
- Mizee, M. R. et al. Isolation of primary microglia from the human post-mortem brain: effects of ante- and post-mortem variables. *Acta Neuropathol. Commun.* **5**, 16 (2017).
- Mildner, A., Huang, H., Radke, J., Stenzel, W. & Priller, J. P2Y<sub>12</sub> receptor is expressed on human microglia under physiological conditions throughout development and is sensitive to neuroinflammatory diseases. *Glia* **65**, 375–387 (2017).
- Moore, C. S. et al. P2Y<sub>12</sub> expression and function in alternatively activated human microglia. *Neurol. Neuroimmunol. Neuroinflamm.* **2**, e80 (2015).
- Gaudilliere, B. et al. Clinical recovery from surgery correlates with single-cell immune signatures. *Sci. Transl. Med.* **6**, 255ra131 (2014).
- Mei, H. E., Leipold, M. D., Schulz, A. R., Chester, C. & Maecker, H. T. Barcoding of live human peripheral blood mononuclear cells for multiplexed mass cytometry. *J. Immunol.* **194**, 2022–2031 (2015).
- Amir, A. D. et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
- Van der Maaten, L. & Hinton, G. Visualizing high-dimensional data using tSNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
- Becher, B. et al. High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.* **15**, 1181–1189 (2014).

28. Holder, G. E. et al. Expression of the mannose receptor CD206 in HIV and SIV encephalitis: a phenotypic switch of brain perivascular macrophages with virus infection. *J. Neuroimmune Pharmacol.* **9**, 716–726 (2014).
29. Cohen, M. et al. Newly formed endothelial cells regulate myeloid cell activity following spinal cord injury via expression of CD200 ligand. *J. Neurosci.* **37**, 972–985 (2017).
30. Roederer, M., Treister, A., Moore, W. & Herzenberg, L. A. Probability binning comparison: a metric for quantitating univariate distribution differences. *Cytometry* **45**, 37–46 (2001).
31. Orlova, D. Y. et al. Earth mover's distance (EMD): a true metric for comparing biomarker expression levels in cell populations. *PLoS One* **11**, e0151859 (2016).
32. Nichols, T. & Hayasaka, S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.* **12**, 419–446 (2003).
33. Chatfield, M. & Mander, A. The Skillings-Mack test (Friedman test when there are missing data). *Stata J.* **9**, 299–305 (2009).
34. Lun, A. T. L., Richard, A. C. & Marioni, J. C. Testing for differential abundance in mass cytometry data. *Nat. Methods* **14**, 707–709 (2017).
35. O'Neill, K., Jalali, A., Aghaeepour, N., Hoos, H. & Brinkman, R. R. Enhanced flowType/RchyOptimix: a BioConductor pipeline for discovery in high-dimensional cytometry data. *Bioinformatics* **30**, 1329–1330 (2014).
36. Smith, A. M. & Dragunow, M. The human side of microglia. *Trends Neurosci.* **37**, 125–135 (2014).
37. Schughart, K., Libert, C. & Kas, M. J. Controlling complexity: the clinical relevance of mouse complex genetics. *Eur. J. Hum. Genet.* **21**, 1191–1196 (2013).
38. Durafourt, B. A., Moore, C. S., Blain, M. & Antel, J. P. Isolating, culturing, and polarizing primary human adult and fetal microglia. *Methods Mol. Biol.* **1041**, 199–211 (2013).
39. Rustenhoven, J. et al. Isolation of highly enriched primary human microglia for functional studies. *Sci. Rep.* **6**, 19371 (2016).
40. Melief, J. et al. Microglia in normal appearing white matter of multiple sclerosis are alerted but immunosuppressed. *Glia* **61**, 1848–1861 (2013).
41. Olah, M. et al. An optimized protocol for the acute isolation of human microglia from autopsy brain samples. *Glia* **60**, 96–111 (2012).
42. Lambert, C., Ase, A. R., Séguéla, P. & Antel, J. P. Distinct migratory and cytokine responses of human microglia and macrophages to ATP. *Brain Behav. Immun.* **24**, 1241–1248 (2010).
43. Klegeris, A., Bissonnette, C. J. & McGeer, P. L. Modulation of human microglia and THP-1 cell toxicity by cytokines endogenous to the nervous system. *Neurobiol. Aging* **26**, 673–682 (2005).
44. Bennett, M. L. et al. New tools for studying microglia in the mouse and human CNS. *Proc. Natl. Acad. Sci. USA* **113**, E1738–E1746 (2016).
45. Bianchin, M. M. et al. Nasu-Hakola disease (polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy–PLOS): a dementia associated with bone cystic lesions. From clinical to genetic and molecular aspects. *Cell. Mol. Neurobiol.* **24**, 1–24 (2004).
46. Korin, B. et al. High-dimensional, single-cell characterization of the brain's immune compartment. *Nat. Neurosci.* **20**, 1300–1309 (2017).
47. Szulzewsky, F. et al. Human glioblastoma-associated microglia/monocytes express a distinct RNA profile compared to human control and murine samples. *Glia* **64**, 1416–1436 (2016).
48. Hamann, J. et al. EMR1, the human homolog of F4/80, is an eosinophil-specific receptor. *Eur. J. Immunol.* **37**, 2797–2802 (2007).
49. Askew, K. et al. Coupled proliferation and apoptosis maintain the rapid turnover of microglia in the adult brain. *Cell Rep.* **18**, 391–405 (2017).
50. Tay, T.L. et al. A new fate mapping system reveals context-dependent random or clonal expansion of microglia. *Nat. Neurosci.* **20**, 793–803 (2017).

## Acknowledgements

We thank C. Böttcher for excellent technical assistance with FACS analysis. We also acknowledge the assistance of the BCRT Flow Cytometry Lab (Charité – Universitätsmedizin Berlin, Germany). C.B. and J.P. were supported by the German Research Foundation (SFB TRR167, B05 & B07). J.P. received additional funding from the Berlin Institute of Health (CRG2aSP6) and the UK DRI (Momentum Award). S.S. was partially funded by the EU-H2020 project PACE (grant agreement number 733006). A.K. and E.P. were supported by stipends from the NeuroMac School (SFB TRR167, IRTG). A.K. received additional funding from the Cluster of Excellence NeuroCure. H.E.M. and A.R.S. were supported through grant Me3644/5-1. B.S. was supported by the German Research Foundation (SI 749/9-1, 749/10-1, CRC-TRR 241). B.S. and R.G. were supported by the Deutsche Krebshilfe (70112011). M.A.M.S. was supported by a 2014 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation and L.D.D.W. by the Virgo Consortium, funded by the Dutch government, project number FES0908. The psychiatric donor program of the Netherlands Brain Bank (NBB-Psy) is financially supported by the Netherlands Organization for Scientific Research (NWO). We acknowledge the Leibniz Science Campus for Chronic Inflammation for general support.

## Author contributions

C.B. and J.P. conceived and designed the project. C.B., S.S., D.K., B.S. and R.G. designed the antibody panels for mass cytometry. M.A.M.S., G.J.L.S., E.M.H., R.S.K. and L.D.D.W. established and performed the isolation of postmortem huMG. P.F. and L.K. provided biopsy tissues from temporal lobe resections. A.R.S. and H.E.M. set up the fixation approach for cryostorage of human leucocytes and provided guidance in using the system. C.B. established the protocol for cryopreservation of isolated huMG. A.K. and E.P. performed barcoding and antibody staining for CyTOF. A.K. conducted FACS analysis of postmortem huMG. D.K. performed CyTOF measurements. E.J.S. and J.P. provided peripheral blood and cerebrospinal fluid samples. C.B. and S.S. analyzed the data. C.B., S.S., L.D.D.W. and J.P. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41593-018-0290-2>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to C.B. or J.P.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

## Methods

**Human blood and CSF samples.** Venous blood and lumbar CSF samples were obtained from four individuals (Supplementary Table 1). The study was registered and approved by the Ethics Commission of Charité–Universitätsmedizin Berlin (registration number EA1/244/12). All study participants provided informed consent before any study-related procedures were undertaken.

**Human brain autopsy.** Human brain tissue was obtained through the Netherlands Brain Bank ([www.brainbank.nl](http://www.brainbank.nl)). The Netherlands Brain Bank received permission to perform autopsies and to use tissue and medical records from the Ethical Committee of the VU University Medical Center. All donors have given informed consent for autopsy and use of their brain tissue for research purposes. Generally, the autopsies of five brain regions (SVZ, THA, CER, GTS and GFM) were performed within 10 h after death. Brain tissue collected for this study was from the donors whose postmortem CSF was between pH 6.5 and 7.2. An overview of the donor information and postmortem variables is summarized in Supplementary Tables 1 and 4.

**Human brain biopsies.** Brain tissue (temporal lobe) was resected for treatment of epilepsy in three patients (Supplementary Table 4) with mesial temporal lobe epilepsy. The resected tissue used in this study was in excess of that needed for pathological diagnosis, and was separated from the epileptic focus with strong epileptogenic activity. Experimental protocol was approved by the Ethics Committee of Charité – Universitätsmedizin Berlin (EA2/111/14) and is in agreement with the Declaration of Helsinki. All patients gave written consent before the surgery.

Specimens were collected in the operating room, transported and processed in cold carboxenated NMDG-aCSF (95% O<sub>2</sub>, 5% CO<sub>2</sub>) containing (in mM): NMDG (93), KCl (2.5), NaH<sub>2</sub>PO<sub>4</sub> (1.2), NaHCO<sub>3</sub> (30), MgSO<sub>4</sub> (10), CaCl<sub>2</sub> (0.5), HEPES (20), glucose (25), sodium L-ascorbate (5), thiourea (2), sodium pyruvate (3).

**Human microglia isolation.** Microglia were isolated according to the previously published protocol<sup>19</sup>. Briefly, the isolation was started within 2 to 25 h after autopsy. Approximately 2–10 g tissue was first mechanically dissociated through a metal sieve in a glucose–potassium–sodium buffer (GKN-BSA; 8 g l<sup>-1</sup> NaCl, 0.4 g l<sup>-1</sup> KCl, 1.77 g l<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 0.69 g l<sup>-1</sup> NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, 2 g l<sup>-1</sup> D-(1)-glucose, 0.3% BSA (Sigma-Aldrich); pH 7.4). For THA (*n* = 8), CER (*n* = 5), GTS (*n* = 18) and GFM (*n* = 16) tissue mixture, the samples were then supplemented with collagenase type I (3,700 units ml<sup>-1</sup>; Worthington) and DNase I (200 µg ml<sup>-1</sup>; Roche Diagnostics GmbH) for 1 h at 37 °C while shaking. For the SVZ samples (*n* = 18), the tissue mixture was subsequently incubated in 0.2% trypsin (Invitrogen) for 20 min at 37 °C. Cell suspension (from all brain regions) was put over a 100 µm cell strainer and washed with GKN-BSA buffer before the pellet was re-suspended in 20 ml GKN-BSA buffer. Next, 10 ml of Percoll (Amersham, GE Healthcare) was added dropwise and tissue was centrifuged at 4,000 r.p.m. for 30 min (4 °C). Three different layers appeared: an upper layer containing myelin, a lower erythrocyte layer and the middle layer containing all cell types, including microglia. The middle layer was carefully taken out without disturbing the myelin layer and washed first with GKN-BSA buffer, followed by magnetic-activated cell sorting buffer (PBS, 1% heat-inactivated FBS (Gibco Life Technologies), 2 mM EDTA). Microglia were positively selected with CD11b-conjugated magnetic microbeads (Miltenyi Biotec GmbH) according to manufacturer's protocol, which resulted in a 99% pure microglia population.

Magnetic-activated cell sorted CD11b<sup>+</sup> (0.5–1 × 10<sup>5</sup>) cells were transferred to a 1.5 ml low-binding Eppendorf (Sigma-Aldrich) and centrifuged at room temperature for 5 min. The cell pellet was then fixed with fixation/stabilization buffer<sup>22</sup> (SmartTube) and frozen at –80 °C until analysis by mass cytometry.

**Flow cytometry.** Cryopreserved microglia (10 GTS, 9 GFM and 10 SVZ; Supplementary Table 4) were thawed and washed twice in staining buffer (PBS containing 0.5% BSA and 2 mM EDTA). Then, cells were stained for CD45 (HI30), CD64 (10.1), CD206 (15-2) and HLA-DR (L-243) in the staining buffer. For intracellular staining, the stained (non-stimulated) cells were then incubated in fixation/permeabilization buffer (eBioscience) for 30 min at 4 °C. Cells were then washed twice with permeabilization buffer (eBioscience). The samples were then stained with anti-CD68 (Y1/82A) antibody permeabilization buffer for 30 min at 4 °C. Cells were subsequently washed once with permeabilization buffer. Forward- and side-scatter parameters were used for exclusion of doublets from analysis. Cellular fluorescence was assessed with CantoII (BD FACSDiva Software 6.1.3; BD Biosciences) and data were analyzed with FlowJo software 10.4.2 (TreeStar) and Cytobank.

**Barcoding. Live cell barcoding.** Individual CSF samples (0.5–1 × 10<sup>4</sup> cells) were pelleted and stained with <sup>89</sup>Y-CD45 (Fluidigm) for 30 min at 4 °C. Cells were then washed and pooled with PBMCs from the same individual.

**Intracellular barcoding.** After fixation and cryopreservation, sorted microglia and CSF-PBMC-pooled samples were thawed and subsequently stained with pre-made combinations of six different palladium isotopes: <sup>102</sup>Pd, <sup>104</sup>Pd, <sup>105</sup>Pd, <sup>106</sup>Pd and <sup>110</sup>Pd (Cell-ID 20-plex Pd Barcoding Kit, Fluidigm). This multiplexing kit applies a 6-choose-3 barcoding scheme that results in 20 different combinations of three

Pd isotopes. After 30 min staining (at room temperature), individual samples were washed twice with cell staining buffer (0.5% BSA in PBS, containing 2 mM EDTA). A total of up to 20 samples (for example, 18 microglia and 2 CSF-PBMC pooled samples), were pooled together, washed and further stained with antibodies. In total, two multiplexed samples (36 microglia and 4 CSF-PBMC pooled samples) were analyzed for microglial regional heterogeneity and phenotypes. For comparative characterization, two multiplexed samples of 19 postmortem huMG (GTS- and GFM-huMG) and three fresh-biopsy-huMG were analyzed.

**Antibodies.** Anti-human antibodies (Supplementary Tables 2 and 3) were purchased either pre-conjugated to metal isotopes (Fluidigm) or from commercial suppliers in purified form and conjugated in-house using the MaxPar X8 kit (Fluidigm) according to the manufacturer's protocol.

**Surface and intracellular staining.** After cell barcoding, washing and pelleting, the combined samples were re-suspended in 100 µl of antibody cocktail against surface markers (Supplementary Tables 2 and 3) and incubated for 30 min at 4 °C. Then, cells were washed twice with cell staining buffer. For intracellular staining, the stained (non-stimulated) cells were then incubated in fixation/permeabilization buffer (eBioscience) for 60 min at 4 °C. Cells were then washed twice with permeabilization buffer (eBioscience). The samples were then stained with antibody cocktails against intracellular molecules (Supplementary Tables 2 and 3) in permeabilization buffer for 1 h at 4 °C. Cells were subsequently washed twice with permeabilization buffer and incubated overnight in 2% methanol-free formaldehyde solution. Fixed cells were then washed, re-suspended in 1 ml iridium intercalator solution (Fluidigm), and incubated for 1 h at room temperature. Next, the samples were washed twice with cell staining buffer and then twice with ddH<sub>2</sub>O (Fluidigm). Cells were pelleted and kept at 4 °C until CyTOF measurement.

**CyTOF measurement.** Cells were analyzed using a CyTOF2 upgraded to Helios specifications, with software version 6.5.236. The instrument was tuned according to the manufacturer's instructions with tuning solution (Fluidigm) and measurement of EQ four-element calibration beads (Fluidigm) containing <sup>140</sup>Ce/<sup>142</sup>Ce, <sup>151</sup>Eu/<sup>153</sup>Eu, <sup>165</sup>Ho and <sup>175</sup>Lu/<sup>176</sup>Lu served as quality control for sensitivity and recovery.

Directly before analysis cells were re-suspended in ddH<sub>2</sub>O, filtered (20 µm Celltrix, Sysmex), counted and adjusted to 3–5 × 10<sup>5</sup> cells ml<sup>-1</sup>. EQ four-element calibration beads were added at a final concentration of 1:10 of the sample volume to be able to normalize the data to compensate for signal drift and day-to-day changes in instrument sensitivity.

Samples were acquired with a flow rate of 300–400 events per second. Lower convolution threshold was set to 400, with noise reduction mode on and cell definition parameters set at an event duration of 10–150 pushes (push = 13 ms). The resulting FCS files were normalized and randomized using the CyTOF software internal FCS-processing module on the non-randomized ('original') data. Settings were used according to the default settings in the software with time interval normalization (100 s per minimum of 50 beads) and passport version 2. Intervals with less than 50 beads per 100 s were excluded from the resulting FCS file.

**Mass cytometry data processing and analysis.** Cytobank was used for initial manual gating on live single cells, Boolean gating for debarcoding and viSNE to generate t-SNE maps<sup>24,25,51</sup>. FCS files containing the t-SNE embedding as additional two parameters were exported from Cytobank for downstream exploratory and statistical analyses using R<sup>52</sup>. The International Society for the Advancement of Cytometry's data standard (Gating-ML 2.0) was used to replicate manual gating within the CytoML/openCyto framework and to update/upload autogenerated gates into Cytobank<sup>53–55</sup>. All FCS files were transformed with Cytobank default arcsinh transformation (scale factor 5). Nucleated single viable cells were manually gated by DNA intercalators <sup>191</sup>Ir/<sup>193</sup>Ir and event length. For debarcoding, Boolean gating was used to deconvolute individual samples according to the barcode combination. For gated circulating cell populations and huMG from different brain regions, expression levels of each marker were assessed and visualized in heat maps. Spearman correlation distance matrices of expression means served as input for Ward's agglomerative hierarchical clustering<sup>56</sup>. Microglia immune phenotypes were visualized using two-dimensional t-SNE maps generated from P2Y<sub>12</sub><sup>+</sup> (Panel A) or TMEM119<sup>+</sup> (Panel B) pre-gated huMG cells. The following 32 markers of Panel A were selected for t-SNE embedding: CCL2, CD115, CD11b, CD11c, CD124, CD16, CD163, CD192, CD195, CD206, CD33, CD37, CD40, CD45, CD56, CD62L, CD64, CD68, CD86, CX3CR1, cyclin A, cyclin B1, EMR1, HLA-DR, IL-10, IRF4, IRF8, Ki-67, P2Y<sub>12</sub>, TGF-β, TNF-α, TREM-2; and 35 markers in Panel B: arginase-1, c-kit, CCL2, CD103, CD116, CD11c, CD135, CD172a, CD18, CD192, CD197, CD206, CD274, CD279, CD32, CD33, CD34, CD36, CD37, CD40, CD44, CD54, CD83, CD86, CD91, C/EBPα, CX3CR1, GM-CSF, HLA-DR, IL-10, IL-6, PU.1, TMEM119, TNF-α, TREM-2. All pre-gated events were used without prior downsampling from 36 samples for each panel (218,986 Panel A, 140,550 Panel B) for embedding using Cytobank's default hyperparameters (perplexity is 30, theta is 0.5 and 1,000 iterations). We found that results were robust in multiple runs as well as to changes of the input parameters; for example, increasing perplexity and number of iterations or removal of some of the markers.

**Statistical t-SNE maps and automated gating of differentially abundant huMG subsets.** As most of the markers exhibited dim or unimodal expression representing a continuum of different phenotypes in the t-SNE maps rather than a number of distinctly segregating clusters of cells, existing methods that rely on clustering techniques<sup>57–59</sup> to first identify a number of subpopulations and then test for differential abundance, for example, could not be applied meaningfully. Therefore, we used a strategy that quantifies and visualizes differences at the level of cellular distribution over t-SNE space<sup>56</sup>. To this end, we generated 2D histograms of the t-SNE maps using the probability binning algorithm available through the R flowFP package<sup>30,60,61</sup>. Dependent on the total number of cells available, a single binning model was created on collapsed data from all samples, by recursively splitting the events at the median values along the two t-SNE dimensions. We chose a grid of 512 bins to have, on average, at least ten cells per bin in each sample for statistical accuracy. Global differences between the cellular density distributions of samples were quantified using the EMD, a histogram similarity metric that also takes the relative location of each bin in the t-SNE map into account. EMD scores were computed (using emdist R package) between each pair of 2D histograms and similarities were visualized by hierarchical clustering in a heat map representation. Regional differences were tested for by permutational analysis of variance using distance matrices<sup>62</sup> as implemented in the function adonis in R package vegan<sup>63</sup>, and setting donor as strata to account for intersubject variability. To identify local features, we performed bin-wise statistical testing for differences in cell frequencies between t-SNE maps from different groups (that is, brain regions)<sup>32,64</sup>. We chose the non-parametric Skillings–Mack<sup>33</sup>, a general Friedman-type statistic to account for the non-normality of cell frequency data and intersubject variability with an incomplete block or repeated measures design (samples from all five brain regions were not available for every donor). Because of the small sample size in this study and many ties (as there usually is a sample-dependent fraction of bins which will not contain any cell), Monte Carlo simulation was used (10,000 permutations) to estimate the null distribution and obtain appropriate Skillings–Mack statistics over the grid of bins. Significant differences are visualized in a statistical t-SNE map where bins are colored on a sliding scale corresponding to their  $-\log_2 P$ , allowing for identification of relevant subsets by spotting areas of connected significant bins<sup>56</sup>. As a complementary approach, we used the R cydar package<sup>34</sup> to detect differentially abundant subsets in original multidimensional marker space. Here, cells are counted into overlapping hyperspheres with a given radius centered at each cell and testing is then performed with the quasi-likelihood method in edgeR<sup>65</sup> and negative binomial generalized linear models to account for overdispersion in the count data. For hypersphere counting on the same markers used for t-SNE embedding we used the default parameterization for the radius, with downsampled data to 20% of total event numbers and a minimum of five cells required to report a hypersphere. Likewise, the result of group-level analysis is visualized on the same composite t-SNE plot for interpreting and exploring significant hyperspheres by coloring their center cells according to  $-\log_2 P$ .

To aid characterization of differentially abundant subsets, compact areas of significant bins (or hyperspheres) are identified using a kernel density-based automated gating algorithm. We used an in-house R implementation which builds on heuristics described by Shekhar et al.<sup>57</sup> for partitioning of t-SNE maps (ACCENSE), adopts methodologies for automated gating of highest density regions in cytometry data<sup>66,67</sup> and utilizes functionalities provided in the R ks package<sup>68,69</sup> for kernel smoothing (source code and description are available in the Supplementary Software or at <https://github.com/steschlick>). A kernel-bandwidth, that is the degree of smoothing, was first chosen to find an estimate of the 2D probability density from the binned data which accurately represented the morphology of the t-SNE map. The  $-\log_2 P$  from bin-wise statistical testing were then integrated as threshold-centered weights into subsequent kernel density estimation after correcting for multiple comparisons by controlling the FDR at 5%<sup>70</sup>. This yielded a smoothed profile of the statistical t-SNE map. Local maxima representing differentially abundant phenotypes were detected by a 2D peak-finding algorithm<sup>57</sup>. Using a standard contouring function, polygons enclosing at least six bins (corresponding to an average cell frequency of 1%) to be reported were grown from each of these peaks to a desired level of significance (Skillings–Mack  $\alpha = 0.005$ ) as tested on the aggregated bins inside a contour. Of note, we also allowed for merging of multiple peaks (or small contours) into a larger single region, as long as these were phenotypically indistinguishable by robust comparison of marker expression profiles as described in the next section.

**Phenotypic characterization and automated annotation of significant subsets.** Identified cellular subsets were automatically labeled with phenotypes according to their distribution of marker expression levels. To describe a subset phenotype, values of a specified  $P_{th}$  upper and lower expression quantile of a given marker were compared with up to three user-defined or data-driven cutoffs for marker positivity. This gave three, six or ten phenotypic categories for a single, two or three cutoffs, respectively. For example, a subset is classified to be positive (+) or negative (–) for a given marker, if more than 84% of cells have expression levels above or below a single cutoff value, respectively. Otherwise this marker is non-informative (0) and will be excluded from the phenotype. Two cutoffs allow phenotypic labeling of subsets for example, defined by a low/dim (+–) or high (+) marker expression. To further facilitate interpretation of subset phenotypes

we quantified differential marker expression between each pair of identified subsets as well as the exclusion of all subset gates (cells–subsets) as an additional reference. Conceptually similar to the marker enrichment modeling described by Diggins et al.<sup>71</sup>, we used a robust effect size, denoted here as  $\Delta$ , that scales the Hodges–Lehmann estimate<sup>72</sup> of the difference in marker expression between two subsets by a robust measure of marker expression variability which we defined as the median absolute deviation about the Hodges–Lehmann estimator<sup>73</sup>. Since manually defined cutoffs might not capture those features, which accounted for the spatial localization of identified subsets in the statistical t-SNE map, computed  $\Delta$  scores were also used to assess the importance of a given marker to algorithmically include only the most relevant markers into data-driven phenotypes that could distinguish the identified huMG subsets. In short, we first filtered out all markers that did not allow discrimination of at least one subset from any other on the basis of a threshold for  $\Delta$  and for the amount of expression overlap; that is, the difference between the  $P_{th}$  lower expression quantile of a subset and upper expression quantile of another subset must be non-negative. For each comparison for which the latter holds, we computed cutoffs discriminating respective subsets and assigned phenotypes accordingly. We set  $P$  to 0.16, corresponding to a  $2\sigma$  difference between two standard normal distributions. This ultimately yields a number of cutoffs per informative marker, ranging from 1 to  $k-1$  for  $k$  subsets.

To reduce redundancy in the phenotype definitions, we used the R lpsolve package to find a combination of at least eight markers that maximized the sum of effective differences  $\Delta$  between any two subsets with one cutoff per marker. This allowed the flowType/RchlyOptimix pipeline<sup>74,75</sup> to be fed with binary gating cutoffs as input so allowing feasible partition of the data into all possible phenotypes and scoring them by the Skillings–Mack test.

**Statistical analysis.** No randomization strategy was used in this study. Data collection and analysis were not performed blind to the conditions of the experiments. However, data processing and analysis were carried out in an unsupervised manner, to exclude the possibility of biased results. CyTOF data are from two multiplexed samples, in which each contains two individual PBMC, two individual CSF cell samples and 18 human brain microglia samples (from up to five brain regions of 4–5 individual donors). No a priori statistical methods were used to predetermine sample sizes due to sample accessibility and insufficient previous data. However, sample sizes were chosen on the basis of estimates of anticipated variability through previous studies of mRNA transcriptomic analysis<sup>16,17</sup>. Quantitative data were shown as independent data points with mean, and analyzed using one-way analysis of variance (ANOVA) with Bonferroni correction for post hoc Tukey multiple comparison testing. Data distribution was assumed to be normal but this was not formally tested. Statistical tests were performed either using GraphPad Prism 6 (GraphPad Software Inc.) or computational analysis using Skillings–Mack non-parametric one-way repeated measures statistic and FDR adjustment for multiple hypothesis testing as described (see Statistical t-SNE maps and automated gating of differentially abundant huMG subsets and Phenotypic characterization and automated annotation of significant subsets). A single GFM sample (from donor number 12) had to be excluded from statistical analyses as the Skillings–Mack testing requires at least two observations (or regions) per block.

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

**Code availability.** The codes used for the data analyses in this study is available in Supplementary Software or at <https://github.com/steschlick>.

## Data availability

Source data associated with Figs. 4–7 can be accessed at <https://flowrepository.org/id/FR-FCM-ZYM6>.

## References

- Kotecha, N., Krutzik, P. O. & Irish, J. M. Web-based analysis and publication of flow cytometry experiments. *Curr. Protoc. Cytom.* **53**, 10.17.1–10.17.24 (2010).
- R Core Team. *R: a language and environment for statistical computing* <http://www.r-project.org/> (R Foundation for Statistical Computing, 2014).
- Finak, G. et al. OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. *PLoS Comput. Biol.* **10**, e1003806 (2014).
- Jiang, M. *CytoML: gatingML interface for openCyto*. R package version 1.0.1 (R Project for Statistical Computing, 2016).
- Spidlen, J., Leif, R. C., Moore, W., Roederer, M. & Brinkman, R. R. Gating-ML: XML-based gating descriptions in flow cytometry. *Cytometry A* **73A**, 1151–1157 (2008).
- van Dongen, S. & Enright, A. J. Metric distances derived from cosine similarity and Pearson and Spearman correlations. Preprint at arXiv <https://arxiv.org/abs/1208.3145> (2012).

57. Shekhar, K., Brodin, P., Davis, M. M. & Chakraborty, A. K. Automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE). *Proc. Natl Acad. Sci. USA* **111**, 202–207 (2014).
58. Chen, H. et al. Cytokit: a Bioconductor package for an integrated mass cytometry data analysis pipeline. *PLoS Comput. Biol.* **12**, e1005112 (2016).
59. Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L. & Nolan, G. P. Automated mapping of phenotype space with single-cell data. *Nat. Methods* **13**, 493–496 (2016).
60. Rogers, W. T. & Holyst, H. A. flowFP: a Bioconductor package for fingerprinting flow cytometric data. *Adv. Bioinforma.* <https://doi.org/10.1155/2009/193947> (2009).
61. Japp, A. S. et al. Wild immunology assessed by multidimensional mass cytometry. *Cytometry A* **91**, 85–95 (2017).
62. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
63. Oksanen, J. et al. *vegan: community ecology package*. (R Project for Statistical Computing, 2008).
64. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* **15**, 1–25 (2002).
65. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
66. Naumann, U., Luta, G. & Wand, M. P. The curvHDR method for gating flow cytometry samples. *BMC Bioinformatics* **11**, 44 (2010).
67. Hahne, F. et al. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**, 106 (2009).
68. Duong, T. ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Stat. Softw.* **21**, 1–16 (2007).
69. Duong, T., Goud, B. & Schauer, K. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proc. Natl Acad. Sci. USA* **109**, 8382–8387 (2012).
70. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
71. Diggins, K. E., Greenplate, A. R., Leelatian, N., Wogslund, C. E. & Irish, J. M. Characterizing cell subsets using marker enrichment modeling. *Nat. Methods* **14**, 275–278 (2017).
72. Hodges, J. L. & Lehmann, E. L. Estimates of location based on rank tests. *Ann. Math. Stat.* **34**, 598–611 (1963).
73. Rousseeuw, P. & Croux, C. Explicit scale estimators with high breakdown point. In *L1-Statistical Analysis and Related Methods* (ed. Dodge, Y.) 77–92 (North-Holland, 1992).
74. Aghaeepour, N. et al. Early immunologic correlates of HIV protection can be identified from computational analysis of complex multivariate T-cell flow cytometry assays. *Bioinformatics* **28**, 1009–1016 (2012).
75. Aghaeepour, N. et al. RchyOptimyx: cellular hierarchy optimization for flow cytometry. *Cytometry A* **81**, 1022–1030 (2012).



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on [statistics for biologists](#) may be useful.

### Software and code

Policy information about [availability of computer code](#)

Data collection

CytoF acquisition software 6.5.236 and flow cytometry acquisition: BD FACSDiva Software 6.1.3

Data analysis

FlowJo software 10.4.2; Cytobank ([www.cytobank.org](http://www.cytobank.org)); R/Bioconductor packages (vegan, emdlist, flowFP, flowWorkspace, CytoML, flowType, RchyOptimyx, edgeR, cydar). Analysis code is available at [https://github.com/steschlick/Boettcher\\_et\\_al\\_NN\\_2018](https://github.com/steschlick/Boettcher_et_al_NN_2018).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data associated with Fig. 4-7 will be available at <https://flowrepository.org/id/FR-FCM-ZYM6>

## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No priori statistical methods were used to predetermine sample sizes due to sample accessibility and insufficient previous data to enable this. However, sample sizes were chosen based on estimates of anticipated variability through previous studies on mRNA transcriptomic analysis (Gosselin et al. 2017; Galatro et al. 2017)
Data exclusions	In Fig. 6b & c and Suppl Fig. 8, one sample (i.e. donor #12 which has a single GFM region) has to be excluded from statistical analyses as testing required at least two observations (i.e. regions) per subject. In Figs 7a & d, two and one sample (out of 36 samples) were excluded from data set of subset 2 and 4, respectively, since these samples contained no cells of these particular subsets. In Supp Fig 10, three, one and six samples (out of 36 samples) were excluded from data set of subset 1, 3 and 4, respectively, since these samples contained no cells of these particular subsets.
Replication	The expression of microglia key markers of each individual samples were measured twice (antibody panel A and B) and were reliably reproduced.
Randomization	no method of randomization was used in this study. Between-subject variability was accounted for with incomplete block and repeated measures designs.
Blinding	No blinding was done, since we performed unsupervised data processing and data analysis, thus excluded the possibility of biased results.

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

CD45 (1:100, HI30 / Fluidigm); CD19 (1:100, HIB19 / Fluidigm); HLA-DR (1:100, L243 / Fluidigm); CD11b (1:100, ICRF44 / Fluidigm); CD124 (1:100, G077F6 / Biolegend); CD64 (1:100, 10.1 / Fluidigm); CD11c (1:100, Bu15 / Fluidigm); CD16 (1:100, 3G8 / Fluidigm); CCL2 (1:200, 5D3-F7 / Biolegend); CD37 (1:100, M-B371 / Biolegend); CD68 (1:100, Y1/82A / Biolegend); TNF- (1:100, Mab11 / Fluidigm); Cyclin B1 (1:100, GNS-1 / Fluidigm); CD3 (1:100, UCHT1 / Fluidigm); CD56 (1:100, B159 / Fluidigm); CCR5 (1:100, NP-6G4 / Fluidigm); Cyclin A (1:100, BF683 / Fluidigm); IRF4 (1:100, 3E4 / Biolegend); CD163 (1:100, GHI/61 / Biolegend); EMR1 (F4/80, A10 / Bio-Rad); Ki-67 (1:100, B56 / Fluidigm); TGF- (1:100, TW4-2F8 / Biolegend); CD115 (1:100, 9-4D2-1E4 / Biolegend); P2Y12 (biotin) (1:100, HPA014518 / Sigma-Aldrich); IL-10 (1:100, JES3-9D7 / Fluidigm); IRF8 (1:100, 7G11A45 / Biolegend); CD206 (1:100, 15-2 / Fluidigm); CD33 (1:100, WM53 / Fluidigm); CD86 (1:100, IT2.2 / Biolegend); CCR2 (1:100, K036C2 / Biolegend); CX3CR (1:100, 12A9-1 / Fluidigm); CD40 (1:50, 5C3 / Biolegend); CD62L (1:100, DREG-56 / Biolegend); CD14 (1:100, M5E2 / Fluidigm); TREM2 (1:100, 237920 / R&D Systems); CD116 (1:100, 4HI / Biolegend); c-kit (1:100, 104D2 / Fluidigm); CD44 (1:50, BJ18 / Biolegend); CD18 (1:100, TS1/18 / Biolegend); CD34 (1:100, 581 / Fluidigm); CD103 (1:100, Ber-ACT8 / Fluidigm); CD83 (1:100, HB15e / Biolegend); IL-6 (1:100, MQ2-13A5 / Biolegend); CD172a (1:100, 15-414 / Biolegend); CD54 (ICAM1) (1:100, HA58 / Biolegend); PD-L1 (1:100, 29E.2A3 / Fluidigm); CD135 (1:100, BV10A4H2 / Fluidigm); GM-CSF (1:100, BVD2-21C11 / Biolegend); CD32 (FITC) (1:100, 6C4 (CD32) / eBioscience); CD36 (1:100, 5-271 / Biolegend); CD91 (1:100, A2MR- 2 (RUO) / BD Bioscience); C/EBP (1:100, 16C12B70 / Biolegend); Arginase-1 (1:100, 658922 / Fluidigm); PU.1 (PE) (1:100, 7C6B05 / Biolegend); CCR7 (1:100, G043H7 / Fluidigm); PD-1 (1:100, EH12.2H7 / Fluidigm); TMEM119 (1:100, HPA052650 / Abcam)

Flow cytometry: CD45 (1:100, HI30, Biolegend), CD64 (1:100, 10.1, Biolegend), CD206 (1:100, 15-2, Biolegend) and HLA-DR (1:100, L-243, Biolegend)

Validation

All antibodies were validated for use in the human myeloid cells using flow cytometry and subsequently mass cytometry.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

In whole, donors included in this study were 8 males (with the age of 39, 46, 48, 57, 61, 69, 72 & 80 years) and 5 females (with the age of 23, 45, 46, 60 (2 donors) & 80 years), see also Supplementary Table 1.

For studies shown in Fig. 3 and 8, the donors were 3 males (with the age of 39, 55 & 69 years) and 10 females (with the age of 23, 39, 45, 46, 55 (2 donors), 60, 74, 88, & 98 years), see also Supplementary Table 4.

Recruitment

Human brain tissue was obtained through the Netherlands Brain Bank ([www.brainbank.nl](http://www.brainbank.nl)). The Netherlands Brain Bank received permission to perform autopsies and to use tissue and medical records from the Ethical Committee of the VU University medical center (VUmc, Amsterdam, The Netherlands). Brain tissue collected for this study was only from the donors whose post-mortem CSF was between pH 6.5 and 7.2.

## Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

see Online Method (p. 28 & 29)

Instrument

see Online Method (p. 29)

Software

see Online Method (p. 29)

Cell population abundance

No cell sorting was performed in this study.

Gating strategy

see Fig. 8 and Supplementary Fig. 1b.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.