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

Chotima Böttcher1,14, Stephan Schlickeiser2,14, Marjolein A M Sneeboer3,4,14, Desiree Kunkel2, Anniki Knop1, Evdokia Paza1,Pawel Fidzinski5, Larissa Kraus5,6, Gijsje J L Snijders3, René S Kahn4, Axel R Schulz7, Henrik E Mei7, NBB-Psy8, Elly M Hol3,9,Britta Siegmund10, Rainer Glauben10, Eike J Spruth1,11, Lot D de Witte3,4,15, Josef Priller1,6,11,12,13,15

1Department of Neuropsychiatry and Laboratory of Molecular Psychiatry, Charité – Universitätsmedizin Berlin, Berlin, Germany. 2Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Berlin, Germany. 3Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht (BCRM-UMCU), Utrecht, The Netherlands. 4Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA. 5Epilepsy-Center Berlin-Brandenburg, Department of Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany. 6 Berlin Institute of Health, Berlin, Germany. 7German Rheumatism Research Center (DRFZ), Berlin, Germany. 8https://www.brainbank.nl/about-us/nbb-psy/. 9Departement of Neuroimmunology, The Netherlands Institute for Neuroscience, An Institute of the Royal Academy of Arts and Sciences, Amsterdam, The Netherlands. 10Medical Department for Gastroenterology, Division of Gastroenterology, Infectiology and Rheumatology, Charité – Universitätsmedizin Berlin, Berlin, Germany. 11DZNE and 12Cluster of Excellence NeuroCure , Berlin, Germany. 13University of Edinburgh and UK Dementia Research Institute, Edinburgh, UK. 14These authors contributed equally to this work. 15These authors jointly directed this work. Correspondence should be addressed to C.B. (chotima.boettcher@charite.de) or J.P. (josef.priller@charite.de).

**ABSTRACT**

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 (103 – 104 cells). We determined expression levels of 57 markers on huMG isolated from up to five different brain regions of nine donors. We identified core 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 permit for the first time to perform high-dimensional, large-scale immunophenotyping of huMG at the single-cell level, which facilitate their unambiguous profiling in health and disease.

**INTRODUCTION**

Microglia are resident innate immune cells of the CNS, and have an important role in maintaining CNS integrity and function1. They are involved in removing apoptotic neurons, refining synaptic connectivity, and providing trophic support for memory and motor learning2,3. In addition, microglia play a role in the development and progression of neurological and psychiatric disorders, including Alzheimer’s disease (AD), amyotrophic lateral sclerosis, schizophrenic psychoses and mood disorders4-7. 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 rodents8-15. 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 microglia16,17. Nonetheless, these studies also demonstrated species-specific patterns of gene expression, and differences in the responses of human and murine microglia to ageing16,17. The heterogeneity of microglia that has been described for the mouse brain7,14 has so far not been replicated in humans. It is also unclear how human and mouse microglia compareon a phenotypic level.

To date, the phenotypic characterization of huMG has mainly relied on immunohistochemical analysis of post-mortem brain tissue, fluorescence-flow cytometric analysis of isolated microglial cells or *in vitro* cultures of huMG18-21. Among the limitations of these approaches are the phenotypic changes induced by cell culture16,19, the high autofluorescent background of post-mortem 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 novel 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 (e.g. cerebrospinal fluid (CSF) and peripheral blood). We isolated huMG from different regions of post-mortem 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 novel protocol for long-term cryopreservation. Subsequently, cryopreserved huMG and peripheral immune cells were simultaneously profiled by multiplexed mass cytometry (CyTOF) using barcoding technology. These powerful methodologies allowed for 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 post-mortem 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 post-mortem brain tissue of different brain regions as described before23. Isolated huMG were then cryopreserved at -80°C using paraformaldehyde-containing stabilizing buffer22. Of note, classical cryopreservation using DMSO failed to cryopreserve the isolated huMG. Up to 3x104 live CSF-derived cells were CD45-barcoded with 89Y-CD45 antibody (CD45-89Y+)23 and subsequently pooled with PBMCs (CD45-89Y-) from the same individual. The CSF-PBMC pooled samples were then cryopreserved using the same protocol that was applied to the huMG. In order 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 5 brain regions of 4-5 donors) (**Supplementary Table 1**). The pooled samples were equally split and stained with two different antibody panels (35 antibodies/panel) (**Supplementary Tables 2 & 3**). *Panel A* was designed to encompass the major circulating immune cell subsets (i.e. T & B cells, monocytes, natural killer (NK) cells) and microglia using proliferation markers, activity-related markers, chemokine receptors and cell subset markers, including P2Y12, IRF4, IRF8, CD45, CD3, CD62L, CD19, HLA-DR, CD56, Cyclin A & B1, Ki67. *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, 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) algorithm24,25 on the commercially available analysis platform Cytobank (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, P2Y12, and based on previous work12,20 was used to identify human microglia (huMG). Notably, P2Y12-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 (e.g. CD11b and CD45) when TMEM119 antibody was used as a huMG marker (*Panel B*, **Supplementary Fig. 1a**). Importantly, virtually all (> 99.9 %) of P2Y12+ cells expressed TMEM119, and > 99.4 % of TMEM119+ cells expressed P2Y12 in the FACS analysis (**Supplementary Figs. 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+ B cells, DCs, CD56dim/+ NK cells, CD3+ T cells) and huMG (P2Y12+ or TMEM119+) clusters on the t-SNE map. Based on the mean signal intensity of all 57 markers analyzed (*Panel A* + *Panel B*), huMG clustered distinctly from all cell subsets in blood and CSF (**Fig. 1c**). Specifically, huMG expressed higher mean levels of P2Y12, TMEM119, EMR1 (F4/80), CD64 and TREM2, whereas expression levels of CD44, CCR2, CD45, CD14 and CD16 were much lower in huMG compared to the PBMCs and CSF cells (**Fig. 1d**). In one huMG sample (out of 36 samples), we could detect an expression of CD19 or CD135 (**Figure 1c**). However, these cells were positive for markers used to identify microglia such as P2Y12, 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 unlikely. Nevertheless, this rare expression of CD19 and CD135 deserves further investigation, which require additional markers. CSF cells expressed comparatively high levels of the cytokines IL-6 and TNF-α (**Supplementary Figs. 2 & 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**). The novel tools allow for the 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 microenvironment26. At the protein expression level, murine microglia can be separated from circulating monocytes and other tissue-resident macrophages by clustering27. Here, we observed that all P2Y12- or TMEM119-expressing huMG also co-expressed HLA-DR and CD11c (**Supplementary Figs. 4a & b**). We therefore performed a comprehensive t-SNE analysis of HLA-DR+CD11c+ huMG and circulating myeloid cells (HLA-DR+CD11c+) 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 then further analyzed these clusters at the level of marker expression on a donor-by-donor, compartment-by-compartment and brain region-by-region basis (**Figs. 2b & c and Supplementary Figs. 2 & 3**). Based on the expression levels of 55 investigated markers (*Panel A* + *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. P2Y12 and TMEM119 expression was highly enriched in huMG and was absent from myeloid cells in blood and CSF (**Figs. 2b & c**).

Next, we compared the immunophenotypic signatures of post-mortem 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, in analogy to the procedure recently describedfor transcriptomic and epigenetic profiling of huMG16. The epileptic focus with the strongest epileptogenic activity was removed before the surrounding tissue was used for microglia isolation. On the t-SNE map, post-mortem huMG clustered together with fresh huMG and displayed a comparable immunophenotypic signature (**Fig. 3a**), underscoring the validity of using post-mortem huMG for immunophenotypic profiling in health and disease. However, we did observe differences in the levels of signal intensity, in particular for IRF8 and P2Y12, and to a lesser degree also for CD11b, CD68 and HLA-DR (**Figs. 3b & c** and **Supplementary Fig. 5a & c**). The expression levels of TMEM119 were not different between post-mortem huMG and fresh huMG (**Fig. 3c**). Interestingly, we detected IRF8hiP2Y12+ cells (G1, **Figs. 3a & b**) at higher frequencies in fresh huMG than in post-mortem huMG (**Fig. 3c**). Removing IRF8 and P2Y12 from embedding parameters resulted in decreased phenotypic differences between post-mortem huMG and fresh huMG (**Figs. 3d & e** and **Supplementary Fig. 5b**), suggesting that IRF8 and P2Y12 are key markers that determine the difference between the two sources of huMG. Of note, these differences may result from post-mortem 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 P2Y12+ cells across different brain regions (**Figs. 4a & b**). CD206 was previously suggested as a marker for M2-macrophagesand perivascular macrophages in the human CNS28. However, recent data obtained from bulk RNA-sequencing showed that huMG also express low levels of CD206 mRNA17,18. Moreover, activated murine microglia express CD206 after spinal cord injury29. 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 P2Y12+ cells derived from up to five brain regions from nine different donors showed a small cluster of CD206highCD163+CD14+ cells (G3), which highly expressed HLA-DR, CD68 and CD11b (**Figs. 4c & d**), suggestive of perivascular macrophages (pmΦ). Interestingly, we observed low expression of P2Y12 on this population (**Figs. 4c & d and Supplementary Fig. 6a**), which has not been reported before12,20. However, the expression levels of P2Y12 were much lower than for huMG (**Supplementary Fig. 6a)**. The cluster of pmΦ 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 CD206lowCD163-CD14-P2Y12high cells (G2 & G3; **Figs. 4c & d**). The CD206low huMG did not express CD163, and showed lower expression levels of HLA-DR, CD68, CD33, CD11b and CD45 compared to pmΦ (**Figs. 4c & d**). However, CD206low huMG showed a higher expression level of CX3CR1 than pmΦ (**Fig. 4d**). In the *Panel B*, the clusters of CD206low huMG were also identified (**Supplementary Figs. 7a & b**). Interestingly, we observed regional heterogeneity in the distribution of CD206low huMG. Quantification of the manually gated populations G1–G3 on the t-SNE display revealed that CD206low huMG were more frequent in temporal lobe (*Gyrus temporalis superior*, GTS) and frontal lobe (*Gyrus frontalis medius*, GFM) compared to other brain regions, whereas CD206high pmΦ were equally distributed across the human brain (**Fig. 4e**). High proportions of CD206low huMG in temporal and frontal lobes were confirmed using *Panel B* (**Supplementary Fig. 7b**). Of note, such low expression of CD206 on P2Y12-expressing cells was failed to be detected by flow cytometry due to the high autofluorescent background of post-mortem huMG (**Supplementary Fig. 6b**). The results underscore the power of the multidimensionality of mass cytometry in the attempt to identify microglia subpopulations in the human brain.

**Heterogeneity of human post-mortem brain microglia**

Recently, region-dependent microglial diversity was detected in the mouse brain based on transcriptional profiling using microarrays14. 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 suggested that huMG from the SVZ display a phenotype that is distinct from huMG in other brain regions (**Figs. 1c & 2b)**. To extend these observations, we performed a comprehensive cluster analysis using the t-SNE embedding of the entire dataset, including all brain regions and all donors (36 samples; **Figs. 5a - c**). In order 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 binning30. The binning model is created on collapsed data from all samples (i.e. concatenated 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 (**Figs. 5d & e**). We used the Earth Mover’s Distance (EMD) metric31 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 (6 of 8 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 Figs. 8a – c**). In particular, donor #6 revealed a distinct cluster of huMG that we further characterized as a CD64hiEMR1hi population (**Supplementary Figs. 8b & c**). Removing this donor or the outlier markers (CD64 and EMR1) prior to t-SNE embedding did not change our results regarding the regional heterogeneity of huMG (**Supplementary Figs. 8d & e)**. We therefore included the outlier donor (#6) and the outlier markers (CD64 and EMR1) in all further analyses so as to embrace the biological variability of huMG.

In order to determine and visualize frequencies of differential phenotypes between brain regions, we performed bin-wise, intra-subject, mass univariate statistical testing32 using the Skillings-Mack (SM) Friedman-type nonparametric one-way repeated measures statistic33 to account for the non-normality of cell frequency data, the incomplete block design (i.e. 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 (**Figs. 6a & b and Supplementary Figs. 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 (subset 1, 2 & 3: *P* < 0.0001; subset 4: *P* = 0.0014, SM-test with controlled false discovery rate (FDR)) (**Figs. 6c - e**). We observed that SVZ and thalamus (THA) contain similar huMG phenotypes (subset 1), which are virtually absent from the other brain regions (**Figs. 6e - g and Supplementary Figs. 8f & g**). Temporal lobe (GTS) and frontal lobe (GFM) are enriched in different huMG phenotypes (subsets 2-4). Interestingly, subset 4 appears to be more abundant in temporal lobe than in frontal lobe (**Figs. 6e - g**). The profile of huMG in the cerebellum (Cer) was distinct from the other brain regions and revealed low abundance of all subsets (**Figs. 6e - g**). Similar regional differences were also detected when antibody *Panel B* was applied to the samples (**Supplementary Figs. 9 & 10**). Finally, we confirmed our findings using the Differential Abundance (DA)-hypersphere analysis in original multi-parameter space with the *cydar/edgeR* framework34 (**Supplementary Figs. 8h & 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 to the other subsets (**Figs. 7a & b** and **Supplementary Figs. 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 to the other subsets (**Figs. 7a & b** and **Supplementary Figs. 10b & d**). Although the two subsets were generally very similar, subsets 2 and 3 differed in their expression of CD64 and EMR1 (**Figs. 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 8 molecules in a reduced binary panel and conventional gating, we applied sequential, i.e. Boolean gating strategies to identify the four putative subsets based on the expression of only these 8 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 (rho/*r2*: subset 1 = 0.87/0.91; subset 2 = 0.85/0.78; subset 3 = 0.87/0.77, **Fig. 7d**), whereas the frequencies of the lower-abundant subset 4 were slightly different between the two types of analysis (rho/*r2*: subset 4 = 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* pipeline35 (**Fig. 7e**). Highly significant 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 *vs* GTS & GFM) based on 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 & 7**) of huMG (**Fig. 8b**, red gate). Importantly, the frequency of this SVZ-enriched subset was comparable between FACS analysis using 4 markers (CD45, CD64, CD68 and HLA-DR) and the CyTOF measurement (**Fig. 8c**).

**DISCUSSION**

Mass cytometry is becoming a known and established

technology enlightening us with the power of multidimen-

sionality w ithin the field of single cell technology. These com-

bined issues of Cytometry Parts A and B demonstrate that the

mass cytometry field is moving forward and palpably contrib-

uting to the overall advancement of cytometry

Mass cytometry is becoming a known and established

technology enlightening us with the power of multidimen-

sionality w ithin the field of single cell technology. These com-

bined issues of Cytometry Parts A and B demonstrate that the

mass cytometry field is moving forward and palpably contrib-

uting to the overall advancement of cytometry

Mass cytometry is becoming a known and established

technology enlightening us with the power of multidimen-

sionality w ithin the field of single cell technology. These com-

bined issues of Cytometry Parts A and B demonstrate that the

mass cytometry field is moving forward and palpably contrib-

uting to the overall advancement of cytometry.

Mass cytometry is becoming a known and established

technology enlightening us with the power of multidimen-

sionality w ithin the field of single cell technology. These com-

bined issues of Cytometry Parts A and B demonstrate that the

mass cytometry field is moving forward and palpably contrib-

uting to the overall advancement of cytometry.

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 animals36,37. 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 cells18,19,38-45. However, phenotypic profiling of huMG based on a comprehensive array of marker proteins has remained technically challenging. This is particularly true at the single-cell level, which is required to identify microglia subpopulations.

Here, we have for the first time 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 microglia14. Our results substantiate previously published data on mRNA16-18,44 and protein18-20,44-46 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 phenotypical signature of huMG that distinguishes them from other innate immune cells (e.g. cells from blood and CSF).

In this study, we combined t-SNE and probability binning for both of their strengths to detect changes in subset 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 8 (out of 57) phenotypical markers that allow to distinguish major huMG subsets.

We confirmed the key transcriptomic signature of huMG16,17 at the protein level, namely, the expression of P2Y12 and TMEM119, the high expression of CD64, CX3CR1, TGF-β1, TREM2, CD115, CCR5, CD32, CD172a, CD91, and the low to absent expression of CD44, CCR2, CD45, CD206, CD163, 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 cells47. These core immunophenotypes of post-mortem huMG are apparently comparable to fresh huMG, albeit differences of signal intensities for some markers. Likewise, Szulzewsky et al.48 observed similar transcriptomic profiles between huMG from epilepsy and postmortem tissues. Interestingly, we detected the expression of EGF-like module containing mucin-like hormone receptor (EMR)1, the human orthologue 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 absent on mononuclear phagocytic cells including monocytes, macrophages and dendritic cell subsets49.

The combination of antibody panel design, single-cell mass cytometry and computational unsupervised data analysis separated cells with characteristics of perivascular macrophages (CD11b+CD206highCD163+) from huMG (CD11b+CD206low/-CD163-). Furthermore, we identified two novel microglial subsets that express CD206, but not CD163, particularly in the frontal and temporal lobe. The findings are in line with the recent studies, which showed low expression of CD206 mRNA in human microglia17,18. Notably, microglia were found to express CD206 after spinal cord injury in mice29, suggesting that CD206 expression may reflect the functional responses of huMG.

A recent study demonstrated that microglia in mice and humans are short lived and quickly renewed at the individual cell level50. At the population level, CNS microglia are maintained by the balanced regulation between proliferation and apoptosis50. In agreement with these observations50, 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 & 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 they reside in, whereas the core immunophenotype that distinguishes them from circulating and/or infiltrating myeloid cells is retained post-mortemly across the five investigated regions of human brain. The results are in line with previous findings in mouse and man14,18,19. 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 to other brain regions. Interestingly, huMG subpopulations in the temporal and frontal lobe expressed 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 determinant.

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 post-mortem 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, Life Sciences Reporting Summary, source data, statements of data availability and associated accession codes are available at the online version of the paper.

**Accession codes**

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

**ACKNOWLEDGEMENTS**

We thank Christian Böttcher for excellent technical assistance with FACS analysis. We would also like to 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 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. was supported by the Deutsche Krebhilfe (70112011). M.A.M.S. was supported by a 2014 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation and L.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.W. established and performed the isolation of post-mortem 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 post-mortem 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.W. and J.P. wrote the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

**REFERENCES**

1. Prinz, M. & Priller, J. Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat. Rev. Neurosci.* **15,** 300–312 (2014).
2. Sierra, A. *et al.* Microglia shape adult hippocampal neurogenesis through apoptosis-coupled phagocytosis. *Cell Stem Cell.* **7**, 483–495 (2010).
3. Parkhurst, C.N. et al. Microglia promote learning-dependent synapse formation through brain-derived neurotrophic factor. *Cell.* **155**, 1596–1609 (2013).
4. Prinz, M. & Priller, J. The role of peripheral immune cells in the CNS in steady state and disease. *Nat. Neurosci.* **20,** 136–144 (2017).
5. Perry, V.H. & Holmes, C. Microglial priming in neurodegenerative disease. *Nat. Rev. Neurol.* **10,** 217–224 (2014).
6. Colonna, M. & Butovsky, O. Microglia function in the central nervous system during health and neurodegeneration. *Annu. Rev. Immunol.* **35,** 441–468 (2017).
7. Keren-Shaul, H. *et al.* A unique microglia type associated with restricting development of Alzheimer’s disease. *Cell* **169,** 1–15 (2017).
8. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330,** 841–845 (2010).
9. Kierdorf, K. *et al.* Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16,** 273–280 (2013).
10. 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).
11. 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).
12. Butovsky, O. *et al.* Identification of a unique TGF-β-dependent molecular and functional signature in microglia. Nat. Neurosci. 17, 131–143 (2014).
13. 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).
14. Grabert, K. *et al.* Microglial brain region−dependent diversity and selective regional sensitivities to aging. Nat. Neurosci. 19, 504–516 (2016).
15. Orre, M. *et al.* Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. *Neurobiol. Aging.* **35,** 2746–2760 (2014).
16. Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science.* **356,** eaal3222 (2017).
17. Galatro, T.F. *et al.* Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat. Neurosci.* **20,** 1162–1171 (2017).
18. Melief, J. *et al.* Characterizing primary human microglia: a comparative study with myeloid subsets and culture models. *Glia* **64,** 1857–1868 (2016).
19. 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).
20. Mildner, A., Huang, H., Radke, J., Stenzel, W. & Priller, J. P2Y12 receptor is expressed on human microglia under physiological conditions throughout development and is sensitive to neuroinflammatory diseases. *Glia* **65,** 375–387 (2017).
21. Moore, C.S. *et al.* P2Y12 expression and function in alternatively activated human microglia. *Neurol. Neuroimmunol. Neuroinflamm.* **2,** e80 (2015)
22. Gaudilliere, B. *et al.* Coordinated surgical immune signatures contain correlates of clinical recovery. *Sci. Transl. Med.* **6**, 255ra131 (2014).
23. Mei, H.E. *et al.* Barcoding of live human peripheral blood mononuclear cells for multiplexed mass cytometry. *J. Immunol.* **194,** 2022–2031 (2015).
24. Amir, E.A. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
25. Van der Maaten, L. & Hinton, G. Visualizing high-dimensional data using tSNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
26. Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
27. 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. *et al.* 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,** 419–46 (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/RchyOptimyx: 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. *et al.* 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. *Gli*a **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. *et al.* Distinct migratory and cytokine responses of human microglia and macrophages to ATP. *Brain Behav. Immun.* **24,** 1241–1248 (2010).
43. Klegeris, A. *et al.* Modulation of human microglia and THP-1 cell toxicity by cytokines endogenous to the nervous system. *Neurobiol. Aging* **26,** 673–682 (2005).
44. Butovsky, O. *et al.* Identification of a unique TGF-β-dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17,** 131–143 (2014).
45. Bennett, M.L. *et al.* New tools for studying microglia in the mouse and human CNS. *Proc. Natl. Acad. Sci. USA* **113**, E1738–1746 (2016).
46. Bianchin, M.M. *et al.* Nasu-Hakola disease (polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy--PLOSL): a dementia associated with bone cystic lesions. From clinical to genetic and molecular aspects. *Cell Mol. Neurobiol.* **24,** 1–24 (2004).
47. Korin, B., *et al.* High-dimensional, single-cell characterization of the brain's immune compartment. *Nat. Neurosci.* **20**, 1300–1309 (2017).
48. Szulzewsky, F., et al. Human glioblastoma-associated microglia/monocytes express a distinct RNA profile compared to human control and murine samples. *Glia*. **64**, 1416–1437 (2016).
49. Hamann, J. *et al.* EMR1, the human homolog of F4/80, is an eosinophil-specific receptor. *Eur. J. Immunol.* **37,** 2797–2802 (2007).
50. Askew, K. et al. Coupled proliferation and apoptosis maintain the rapid turnover of microglia in the adult brain. Cell Rep. 18, 391–405 (2017).

**FIGURE LEGENDS**

**Figure 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 cerebrospinal fluid (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 P2Y12 (red denotes high expression, blue denotes no expression). P2Y12+ cells were gated as huMG (green) and P2Y12- cells were gated as different circulating immune cells. (**c**) Heat map and cluster analysis of all samples based on the mean expression of 57 markers (*Panel A* + *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) = subventricular zone; THA (grey) = thalamus; CER (red) = cerebellum; temporal lobe (GTS, bright blue); frontal lobe (GFM, dark blue) samples and expression levels are indicated by dendrograms. Heat colours of expression levels have been scaled for each marker individually (to the 1st and 5th quintiles), while red denotes high and blue low expression. (**d**) Mean signal intensity levels of P2Y12, 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 data sets).

**Figure 2** Comparatively phenotypic analysis of huMG and peripheral myeloid cells. (**a**) Representative t-SNE projection of multi-dimensional single cell phenotypes of HLA-DR+CD11c+ 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 colour spectrum represents an expression level of P2Y12 (red = high; dark blue = no expression). The bottom image showed 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* + *Panel B*) and relationships between blood (blue), CSF (orange) and brain myeloid cells (SVZ = bright green; THA = grey; CER = red; GTS = bright blue; GFM = dark blue). Heat colours have been scaled per marker (red denotes high and blue denotes low expression). (**c**) Mean expression levels of selected markers in blood, CSF and brain myeloid cells (black lines show mean values of the data sets).

**Figure 3** Comparative immune profiling of post-mortem huMG and fresh huMG. (**a**) An overlaid t-SNE plot of all cells from all samples (green = post-mortem GTS-huMG; n = 10 biologically independent samples; orange = post-mortem GFM-huMG, n = 9 biologically independent samples and 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 samples (black lines show mean values of the data sets). \*\*\*\**P* <0.0001, one-way ANOVA with Bonferroni correction. (**b**) Representative reduced-dimensional single cell t-SNE illustrations of P2Y12+ huMG from biologically independent samples of GTS (n = 10), GFM (n = 9) and fresh biopsies (n = 3). The colour spectrum represents an expression level (red = high; dark blue = low expression). (**c**) Mean expression levels of selected markers showing differential marker expressions between the two gates (G1 & 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 P2Y12 and IRF8) of all cells from all biologically independent samples (green = GTS, n = 10; orange = GFM, n = 9 and blue = fresh biopsies, n = 3). No distinct cluster was detected, thus minute differences between samples. (**e**) Representative reduced-dimensional single cell t-SNE illustrations of P2Y12+ huMG from biologically independent samples of GTS (n = 10), GFM (n = 9) and fresh biopsies (n = 3). The colour spectrum represents an expression level of CD45, HLA-DR, CD11b and CD68 (red = high; dark blue = low expression).

**Figure 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 P2Y12 and CD206. (**b**) Mean expression level of CD206 in P2Y12+ 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 data sets. (**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, P2Y12, HLA-DR, CD163, CD11b and CD68. Red colour denotes high expression, blue colour denotes no expression. CD206high perivascular mΦ is gated as “G3” (red square), and CD206low cell population is gated as “G1” (orange circle) and “G2” (green circle). (**d**) Histogram plots show an expression of selected markers in CD206-negative (blue), CD206low (G1, orange & G2, green) huMG and CD206high 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 represents the mean value.

**Figure 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. The coloring indicates (**a**) nine donors, (**b**) five brain regions, or (**c**) overall cellular density of the concatenated files (spectrum from blue (low density) to red (high density)). (**d**) From the concatenated files a single binning grid is established, comprising 512 “micro-gates” in which cell frequencies are enumerated. Superimposing the binning grid on individual samples’ t-SNE landscapes allows comparative analysis of huMG profiles at sample-to-sample basis across all five regions. (**e**) t-SNE plots of concatenated FCS files of each brain region. (**f**) Heatmap representing the pairwise earth mover’s distances (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).

**Figure 6** Region-dependent huMG-subpopulations. (**a**) Binning grid superimposed on concatenated t-SNE map (n = 36 biologically independent samples) and (**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) and GFM (n = 6)) using the nonparametric Skillings-Mack statistic for unbalanced two-way block designs. The colour spectrum corresponds to unadjusted *P*-values, ranging from <0.000001 (red) to 1 (blue). (**c**) Smoothed representation of statistical t-SNE map (**Fig. 6b**) 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). Gates of four detected subsets (black lines) are shown in concatenated t-SNE maps, color-encoded to indicate (**d**) nine donors, (**e**) five brain regions, and (**f**) in concatenated cellular density plots for each of the five investigated brain regions, 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 and GFM, n = 7). Black lines show mean of the data set. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* <0.0001, one-way ANOVA with Bonferroni correction.

**Figure 7** Region-dependent huMG phenotypes. (**a**) Radar chart (or snail plot) shows marker expression levels of each huMG subset (subset 1 = red; 2 = orange; 3 = green and 4 = purple). The snail shell represents transversal (perpendicular) axis mapping marker expression levels on exponential scale. Each line denotes each sample (n = 36 biologically independent samples). The right image demonstrates the automated subset gating on t-SNE map (subset 1 = red; 2 = orange; 3 = green; 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 boxplot representation (subset 1 = red box (n = 66 bins); 2 = orange box plot (n = 32 bins); 3 = green box (n = 33 bins); 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 1.5x 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 and GFM, n = 7) are overlaid with the four subsets identified by multivariate gating using a (Boolean) combination of 1D-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, *r2* 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 and GFM, n = 6) from 8 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+ subset 3 phenotype has been included to also target the fraction of cells with CD206-expression above cutoff.

**Figure 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 that 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 10 donors). \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* <0.0001, one-way ANOVA with Bonferroni correction. (**b**) Representative multi-dimensional single cell t-SNE projections of huMG from biologically independent samples of SVZ (n = 10), GTS (n = 10) and GFM (n = 9). The colour spectrum represents an expression levels of CD45, CD68, HLA-DR and CD64 (red = high; dark blue = low expression). The left image shows overlaid multi-dimensional plot of all cells from all three brain regions (green = SVZ, blue = GTS and 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 technique (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 data sets.

ONLINE METHODS

**Human blood and CSF samples.** Venous blood and lumbar cerebrospinal fluid (CSF) samples were obtained from four individuals (**Supplementary Table 1**). The study was registered and approved by the Ethics Commission of Charité–Universitätsmedizin Berlin (Ethikkommission der Charité–Universitätsmedizin Berlin; registration number EA1/244/12), Berlin, Germany. 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 (VUmc, Amsterdam, The Netherlands). All donors have given informed consent for autopsy and use of their brain tissue for research purposes. Generally, the autopsies of five brain regions, which were subventricular zone (SVZ), thalamus (THA), cerebellum (CER), temporal lobe (*Gyrus temporalis superior*, GTS) and frontal lobe (*Gyrus frontalis medius*, GFM), were performed within 10h after death. Brain tissue collected for this study was from the donors whose post-mortem CSF was between pH 6.5 and 7.2. An overview of the donor information and post-mortem variables is summarized in **Supplementary Table 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 (mTLE). The resected tissue used in this study was in excess of that was 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 prior to the surgery.

Specimens were collected in the operation room, transported and processed in cold carbogenated NMDG-aCSF (95% O2, 5% CO2) containing (in mM): NMDG (93), KCl (2.5), NaH2PO4 (1.2), NaHCO3 (30), MgSO4 (10), CaCl2 (0.5), HEPES (20), glucose (25), Na-L-ascorbate (5), thiourea (2), Na-pyruvate (3).

**Human microglia isolation.** Microglia was isolated according to the previously published protocol19. Briefly, the isolation was started within 2 to 25 hours after autopsy. Approximately, 2-10 grams tissue was first mechanically dissociated through a metal sieve in a glucose-potassium-sodium buffer (GKN-BSA; 8g/l NaCl, 0.4g/l KCl, 1.77g/l Na2HPO4.2H2O, 0.69 g/l NaH2PO4.H2O, 2 g/l D-(1)- glucose, 0.3% bovine serum albumin (BSA, Sigma-Aldrich); pH 7.4). For THA (n = 8), CER (n = 5), GTS (n = 18) & GFM (n = 16) tissue mixture, the samples were then supplemented with collagenase Type I (3700 units/ ml; Worthington, Lakewood, NJ, USA) and DNase I (200 µg/ml; Roche Diagnostics GmbH) for 1 h at 37°C while shaking. For the SVZ (n = 18) samples, the tissue mixture was subsequently incubated in 2.5% trypsin (Invitrogen) for 20 min at 37°C. Cell suspension (from all brain regions) was put over a 100µM cell strainer and washed with GKN-BSA buffer before the pallet was re-suspended in 20ml GKN-BSA buffer. Next, 10ml of Percoll (Amersham, GE Healthcare) was added dropwise and tissue was centrifuged at 4000 rpm for 30 minutes (4°C). Three different layers appeared: 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 (MACS) buffer (PBS, 1% heat-inactivated fetal calf serum (FCS, Gibco Life technologies, Massachusetts), 2mM 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.

MACS-sorted CD11b+ (0.5 – 1x105) cells were transferred to a 1.5ml low binding Eppendorf (Sigma-Aldrich, the Netherlands) and centrifuged at room temperature (RT) for 5 minutes. Cell pallet was then fixed with fixation/stabilization buffer22 (SmartTube) and frozen at –80°C until analysis by mass cytometry.

**Flow cytometry.**Cryopreserved microglia (10 GTS; 9 GFM & 10 SVZ; **Supplementary Table 4**) were thawed and washed twice in staining buffer (PBS containing 0.5% bovine serum albumin and 2mM 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 (Fix/Perm Buffer, eBioscience) for 30 min at 4°C. Cells were then wash 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 – 1x104 cells) were pelleted and stained with 89Y-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 premade combinations of six different palladium isotopes: 102Pd, 104Pd, 105Pd, 106Pd, 108Pd & 110Pd (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 RT), individual samples were washed twice with cell staining buffer (0.5% bovine serum albumin in PBS, containing 2mM EDTA). Total of up to 20 samples (e.g. 18 microglia & 2 CSF-PBMC pooled samples), were pooled together, washed and further stained with antibodies. In total of two multiplexed samples (36 microglia & 4 CSF-PBMC pooled samples) were analysed for microglial regional heterogeneity. For a comparative characterization, two multiplexed samples of 19 post-mortem-huMG (GTS- and GFM-huMG) and 3 fresh-biopsy-huMG) were analysed.

**Antibodies.** Anti-human antibodies (**Supplementary Table 2 & 3**) were purchased either preconjugated 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 Table 2 & 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 (Fix/Perm Buffer, eBioscience) for 60 min at 4°C. Cells were then wash twice with permeabilization buffer (eBioscience). The samples were then stained with antibody cocktails against intracellular molecules (**Supplementary Table 2 & 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 (FA). Fixed cells were then washed and re-suspended in 1 ml iridium intercalator solution (Fluidigm) for 1 h at RT.Next, the samples were washed twice with cell staining buffer and then twice with ddH2O (Fluidigm). Cells were pelleted and kept at 4°C until CyTOF measurement.

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

Directly prior to analysis cells were re-suspended in ddH2O, filtered (20 µm Celltrix, Sysmex), counted and adjusted to 3–5x105 cells/ml. 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/second. Lower convolution threshold was set to 400, with noise reduction mode on and cell definition parameters set at event duration of 10-150. The resulting flow cytometry standard (FCS) files were normalized and randomized using the CyTOF software's 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 seconds/minimum of 50 beads) and passport version 2. Intervals with less than 50 beads per 100 seconds 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 maps51-53. FCS files containing the t-SNE embedding as additional two parameters were exported from Cytobank for downstream exploratory and statistical analyses using R54. 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 auto-generated gates into Cytobank55-57. All FCS files were transformed with Cytobank default arcsinh transformation (scale factor 5). Nucleated single viable cells were manually gated by DNA intercalators 191Ir/193Ir and event length. For debarcoding, Boolean gating was used to deconvolute individual sample 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 heatmaps. Spearman correlation distance matrices of expression means served as input for Wards agglomerative hierarchical clustering58. Microglia immune phenotypes were visualized using two- dimensional t-SNE maps generated from P2Y12+ (*Panel A*) or TMEM119+ (*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, P2Y12, TGF-b, 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/EBPa, CX3CR1, GM-CSF, HLA-DR, IL-10, IL-6, PU.1, TMEM119, TNF-a, TREM-2. All pre-gated events were used without prior downsampling from 36 samples for each panel (218986 *Panel A*, 140550 *Panel B*) for embedding using Cytobank’s default hyperparameters (perplexity=30, theta=0.5, and 1000 iterations). We found that results were robust in multiple runs as well as to changes of the input parameters, e.g. 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 in a continuum of different phenotypes in the tSNE maps rather than a number of distinctly segregating clusters of cells, existing methods that rely on clustering techniques59-61 to first identify a number of subpopulations and then test for e.g. differential abundance could not meaningfully be applied. Therefore, we used a strategy that quantifies and visualizes differences at the level of cellular distribution over t-SNE space58. To this end, we generated 2D-histograms of the t-SNE maps using the probability binning algorithm available through the R flowFP package62-64. 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 10 cells per bin in each sample for statistical accuracy. Global differences between samples’ cellular density distributions were quantified using the earth mover's distance (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 heatmap representation. Regional differences were tested for by permutational analysis of variance using distance matrices65 as implemented in the function adonis in R package vegan66 and setting donor as strata to account for inter-subject variability. In order to identify local features, we performed bin-wise statistical testing for differences in cell frequencies between t-SNE maps from different groups, i.e. brain regions67,68. We chose the nonparametric Skillings-Mack69 (SM), a general Friedman-type statistic to account for the non-normality of cell frequency data and inter-subject variability with an incomplete block or repeated measures design (i.e. 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 SM statistics over the grid of bins. Significant differences are visualized in a statistical tSNE map where bins are coloured on a sliding scale corresponding to their –log2 *P*-values, allowing for identification of relevant subsets by spotting areas of connected significant bins58. As a complementary approach, we utilized the R cydar package70 to detect differentially abundant subsets in original multidimensional marker space. Here, cells are counted into overlapping hyperspheres with a given radius centred at each cell and testing is then performed with the quasi-likelihood method in edgeR71 and negative binomial generalized linear models to account for overdispersion in the count data. For hypersphere counting on the same markers used for tSNE embedding we used the default parametrization for the radius, with downsampled data to 20 percent of total event numbers and a minimum of 5 cells required to report a hypersphere. Likewise, the result of group-level analysis is visualized on the same composite tSNE plot for interpreting and exploring significant hyperspheres by colouring their centre cells according to –log2 *P*-values.

In order 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 upon heuristics described by Shekhar *et al*.59 for partitioning of tSNE maps (ACCENSE), adopts methodologies for automated gating of highest density regions in cytometry data72,73 and utilizes functionalities provided in the R ks package74,75 for kernel smoothing (source code and description are available in the Supplementary Software or at https://github.com/steschlick). A kernel-bandwidth, i.e. 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 –log2 *P*-values from bin-wise statistical testing were then integrated as threshold-centred weights into subsequent kernel density estimation after correcting for multiple comparisons by controlling the false discovery rate (FDR) at 5%76. This yielded a smoothed profile of the statistical tSNE map. Local maxima representing differentially abundant phenotypes were detected by a 2D peak-finding algorithm59. Using a standard contouring function, polygons enclosing at least 6 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 (SM α=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 labelled with phenotypes according to their distribution of marker expression levels. In order to describe a subset phenotype, values of a specified pth 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 3, 6, or 10 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 labelling of subsets e.g. 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.*77, we used a robust effect size, denoted here as Δ, that scales the Hodges-Lehmann estimate78 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 estimator79. 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 Δ scores were also used to assess the importance of a given marker in order 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 to discriminate at least one subset from any other based on a threshold for Δ and for the amount of expression overlap, i.e. the difference between the pth 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 for k subsets from 1 to k-1.

In order to reduce redundancy in the phenotype definitions, we used the R lpsolve package to find a combination of at least 8 markers that maximized the sum of effective differences Δ between any two subsets with one cutoff per marker. This allowed to fed the *flowType/RchyOptimyx pipeline*80,81 with binary gating-cutoffs as input to feasibly partition the data into all possible phenotypes and score them by SM 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 eighteen human brain microglia samples (from up to five brain regions of 4- 5 individual donors). No priori statistical methods were used to predetermine sample sizes due to sample accessibility and insufficient previous data. However, sample sizes were chosen based on estimates of anticipated variability through previous studies of mRNA transcriptomic analysis16,17. Quantitative data were shown as independent data points with mean, and analyzed using one-way ANOVA with Bonferroni correction for post-hoc Turkey 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 Skilling-Mack nonparametric one-way repeated measures statistic and FDR-adjustment for multiple hypothesis testing as described above (*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 #12) has to be excluded from statistical analyses as the Skilling-Mack testing requires at least two observations (i.e. regions) per block.

**Data availability**

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

**Code availability**

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

**Reference**

1. 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).
2. Amir el, A.D., *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. biotech.* **31**, 545–552 (2013).
3. van der Maaten, L.J.P. & Hinton, G.E. Visualizing High-Dimensional Data Using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
4. R\_Core\_Team *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL [http://www.r-project.org/](http://www.R-project.org/) (2014).
5. 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).
6. Jiang, M. *CytoML: GatingML interface for openCyto.* R package version 1.0.1 (2016).
7. Spidlen, J. et al. Gating-ML: XML-based gating descriptions in flow cytometry. *Cytometry A.* **73A**, 1151–1157 (2008).
8. van Dongen, S. & A.J. Enright. *Metric distances derived from cosine similarity and Pearson and Spearman correlations.* CoRR, arXiv:1208.3145 [stat.ME] (2012).
9. Shekhar, K., Brodin, P., Davis, M.M. & Chakraborty, A.K. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). *Proc. Nat. Acad. Sci. USA.* **111**, 202–207 (2014).
10. Chen, H. *et al*. Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline. *PLoS Comput. Biol.* **12**, e1005112 (2016).
11. 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).
12. Rogers, W.T. & Holyst, H.A. FlowFP: A Bioconductor Package for Fingerprinting Flow Cytometric Data. *Adv. Bioinformatics.* 193947 (2009).
13. Roederer, M. *et al.* Probability binning comparison: a metric for quantitating univariate distribution differences. *Cytometry*. **45,** 37–46 (2001).
14. Japp, A.S., *et al.* Wild immunology assessed by multidimensional mass cytometry. Cytometry A. 91, 85–95 (2017).
15. Anderson, M.J. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26, 32–46 (2001).
16. Oksanen, J., *et al.* Vegan: *Community Ecology Package.* (2008).
17. Nichols, T. & Hayasaka, S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.* **12,** 419 – 446 (2003).
18. Nichols, T.E. & Holmes, A.P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1 – 25 (2002).
19. Chatfield, M. & Mander, A. The Skillings-Mack test (Friedman test when there are missing data). *Stata J.* **9**, 299–305 (2009).
20. Lun, A.T.L., Richard, A.C. & Marioni, J.C. Testing for differential abundance in mass cytometry data. *Nat. Methods.* **14**, 707–709 (2017).
21. 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).
22. Naumann, U., Luta, G. & Wand, M.P. The curvHDR method for gating flow cytometry samples. *BMC Bioinformatics* **11**, 44 (2010).
23. Hahne, F., *et al.* flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**, 106 (2009).
24. Duong, T. ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R. *J. Stat. Softw.* **21,** 1–16 (2007).
25. 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).
26. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R Stat. Soc. Ser. B.* **57,** 289–300 (1995).
27. Diggins, K.E. *et al.* Characterizing cell subsets using marker enrichment modeling. *Nat. Methods.* **14,** 275–278 (2017).
28. Hodges, J.L. & Lehmann, E.L. Estimates of Location Based on Rank Tests. *Ann. Math. Statist.* **34**, 598–611 (1963).
29. Rousseeuw, P. & Croux, C. *Explicit Scale Estimators with High Breakdown point* (1992).
30. 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).
31. Aghaeepour, N. *et al.* RchyOptimyx: cellular hierarchy optimization for flow cytometry. *Cytometry A.* **81,** 1022–1030 (2012).