Article

Unraveling microglial spatial organization in the developing human brain with DeepCellMap, a deep learning approach coupled with spatial statistics

Theo Perochon[1], Zeljka Krsnik [2], Marco Massimo[3], Yana Ruchiy [4], Alejandro Lastra Romero[4], Elyas Mohammadi[5], Xiaofei Li [5], Katherine R. Long[3], Laura Parkkinen [6], Klas Blomgren [4], Thibault Lagache [7,9] ✉, David A. Menassa [4,6,9] ✉ & David Holcman [1,8,9] ✉

Mapping cellular organization in the developing brain presents significant challenges due to the multidimensional nature of the data, characterized by complex spatial patterns that are difficult to interpret without high-throughput tools. Here, we present DeepCellMap, a deep-learning-assisted tool that integrates multi-scale image processing with advanced spatial and clustering statistics. This pipeline is designed to map microglial organization during normal and pathological brain development and has the potential to be adapted to any cell type. Using DeepCellMap, we capture the morphological diversity of microglia, identify strong coupling between proliferative and phagocytic phenotypes, and show that distinct spatial clusters rarely overlap as human brain development progresses. Additionally, we uncover an association between microglia and blood vessels in fetal brains exposed to maternal SARS-CoV-2. These findings offer insights into whether various microglial phenotypes form networks in the developing brain to occupy space, and in conditions involving haemorrhages, whether microglia respond to, or influence changes in blood vessel integrity. DeepCellMap is available as an open-source software and is a powerful tool for extracting spatial statistics and analyzing cellular organization in large tissue sections, accommodating various imaging modalities. This platform opens new avenues for studying brain development and related pathologies.

Recent advancements in digital pathology and multi-color fluorescence microscopy, including traditional techniques such as haematoxylin and eosin (H&E) staining[1] and immunohistochemistry[2], as well as more sophisticated multiplex imaging approaches[3,4], have transformed our ability to map cellular organization directly within tissues. Coupled with automated image analysis, these methods have provided invaluable insights into key biological processes, spanning fields such as cancer biology, immunology, and developmental neuroscience. While fluorescence and brightfield microscopy have been widely employed, their potential is enhanced when integrated with advanced spatiotemporal cell mapping techniques, which enable the comprehensive analysis of cellular organization, dynamics, and morphology in complex tissue environments.

A full list of affiliations appears at the end of the paper. ✉e-mail: thibault.lagache@pasteur.fr; david.menassa@queens.ox.ac.uk; david.holcman@ens.psl.eu

Microglial cells, originating from extraembryonic yolk sac progenitors, migrate into the developing brain from the 4th post-conceptional week (pcw) and colonize the telencephalon through migration and cycles of proliferation and apoptosis[5]. During this process, microglia display marked morphological heterogeneity, reflective of their diverse roles in brain development and neurodevelopmental topography establishment[5–10]. Understanding how various microglial morphologies are organized within the developing brain and the spatial relationships between their distinct morphologies can provide critical insights into their functional roles. Despite their importance, there is a lack of high-throughput, automated tools capable of mapping these interactions, limiting our ability to fully characterize microglial organization during brain development, particularly in humans. Microglial organization is known to be disrupted in various pathological conditions[11,12]. For instance, hypoxic conditions drive microglial proliferation, leading to altered spatial distributions and cellular organization[13]. In addition, maternal infection with SARS-CoV-2 during pregnancy has been linked to fetal cortical hemorrhages, where compromised blood vessel integrity may disrupt microglial organization[14]. However, it remains unclear to what extent microglial spatial dynamics are affected under these pathological conditions. Understanding how microglia respond to such perturbations may help elucidate their role in injury and repair mechanisms, particularly in relation to blood vessel integrity.

Historically, the study of microglial morphology has been largely descriptive, relying on manual quantification methods that are time-intensive and prone to bias[5,7]. Automating the recognition of microglial morphology and extracting spatial statistics to analyze their interrelationships allow for uncovering organizational patterns that are otherwise missed by manual approaches. Recent developments in automated image analysis, particularly with the advent of deep learning algorithms, have significantly improved the segmentation of cells and tissue regions[15,16]. Both supervised and unsupervised machine learning models have emerged as effective tools for whole-slide detection of microglia, particularly in brightfield microscopy, where segmentation accuracy is critical. While these advances have shown great promise in mouse brain sections, their application to human post-mortem tissues is still in the early stages[17,18].

Accurate cell segmentation is a critical first step in the meaningful analysis of tissue organization. In fluorescence microscopy, where cell types are often color-coded, segmentation alone can reveal cellular distributions. However, in brightfield microscopy, an additional classification step is required, which must consider both morphological features and spatial information related to the cell's local neighborhood. This is particularly important in developing tissues, where various cell phenotypes, such as microglial cells, frequently intermingle. Current methods, primarily designed for postnatal non-human tissues, face scalability challenges due to small sample sizes and reliance on manual segmentation. These approaches often struggle with the complexity and diversity of cell morphologies encountered during development, where there is significant interplay between different cell types[9,19].

To gain deeper insights into the spatial relationships between distinct morphologies[20,20], advanced spatial statistics can be leveraged to uncover associations between different cell phenotypes and to analyze their spatial distributions within the tissue. Traditionally, spatial analyses in digital pathology have been limited to methods that quantify the accumulation of specific cell types in regions (e.g., tumor islets)[21], or that evaluate cellular neighborhoods based on local cell type stoichiometry[22]. However, deeper spatial statistics that examine cell type proximity (called cell-to-cell association or coupling see Supplementary Material for specific definitions)[23] remain under utilized[24]. These frameworks are often difficult to generalize to complex cell shapes, like microglia, and to quantify the accumulation of cells at an anatomical region's boundary (that we call cell-region association, see Supplementary Material). Moreover, these statistics

require careful tissue-specific considerations (e.g., cell density, regional boundaries) and statistical processing to ensure accurate biological interpretation[25].

A promising approach to characterize cell-to-cell and cell-to-region spatial relationships involves a recent methodology called levelset analysis that partitions the tissue into sub-regions with respect to the distance to a given cell type or region. Then, the statistical analysis of the number of a second population of cells (e.g., another morphological type of microglia) in each sub-region enables the robust characterization of how certain cell types accumulate around others or in proximity to specific tissue regions[25]. However, such analysis must also account for uncertainties arising from the deep learning-based classification of cells.

In this work, we introduce DeepCellMap, a computational approach compiled into an open-source Python package developed for robust morphological cell classification using deep learning, coupled with advanced statistical characterization of spatial relationships between different cell morphologies/classes and their distribution within tissues. DeepCellMap integrates a generalized Ripley's method that accounts for classification uncertainties in the deep neural networks, enabling a detailed analysis of cell distributions in tissues and with respect to the boundary of a brain region. To further investigate cell clusters, their organization, and the spatial dynamics of mixed microglial phenotypes, this method also features an optimized Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm[26], with a semi-automated parameter calibration. Compatible with both brightfield and fluorescent microscopy images, we apply DeepCellMap to histological brain images that we labeled with immunohistochemistry and immunofluorescence from normal human fetuses and fetuses whose mothers were exposed to SARS-CoV-2 during pregnancy[14,27]. By fusing machine learning with advanced spatiotemporal statistics, this approach enables the characterization of microglial organization over time in both healthy and diseased conditions. The flexibility of this method also allows for its adaptation to diverse datasets of dynamically changing tissues, offering a powerful tool for broad applications in biomedical and clinical research.

Results
Cell morphology classification and advanced spatial statistics with DeepCellMap
Cell segmentation. To map the distribution and spatial relationships between various cell classes or morphologies in tissue samples, derived from different imaging modalities such as brightfield immunohistochemistry or high-resolution confocal microscopy, we developed DeepCellMap, a fully automated analysis pipeline processes labeled images as input (Fig. 1A and Supplementary Fig. 1A). Given the typically large size of these images (many Gigabytes), a pre-segmentation step divides them into smaller, manageable tiles (Supplementary Figs. 2 and 3). In each of these tiles, cell shapes are automatically segmented (Supplementary Fig. 1B1) using image thresholding and morphological operations, ensuring the precise identification of cellular boundaries and features (Supplementary Fig. 4 for method and Fig. 5 for validation).

Cell classification. Microglial cells display an array of morphologies in homeostatic, pathological, and developmental conditions in humans and rodents. Microglia have been described as amoeboid, migratory, rod-shaped, phagocytic, ramified, intermediate/hyper-ramified, multinucleated, activated, reactive, dark, dystrophic, aggregated, bulbous, and more[5,7,28–33]. Some of these morphologies evolve from one to the next across developing tissues and are tied to function, as shown in Supplementary Fig. 6A. Some morphologies coupled with functional markers inform us about what these cells are doing such as proliferative, dying and phagocytic microglia. To detect these cells, position, local neighborhood and shape are key parameters.
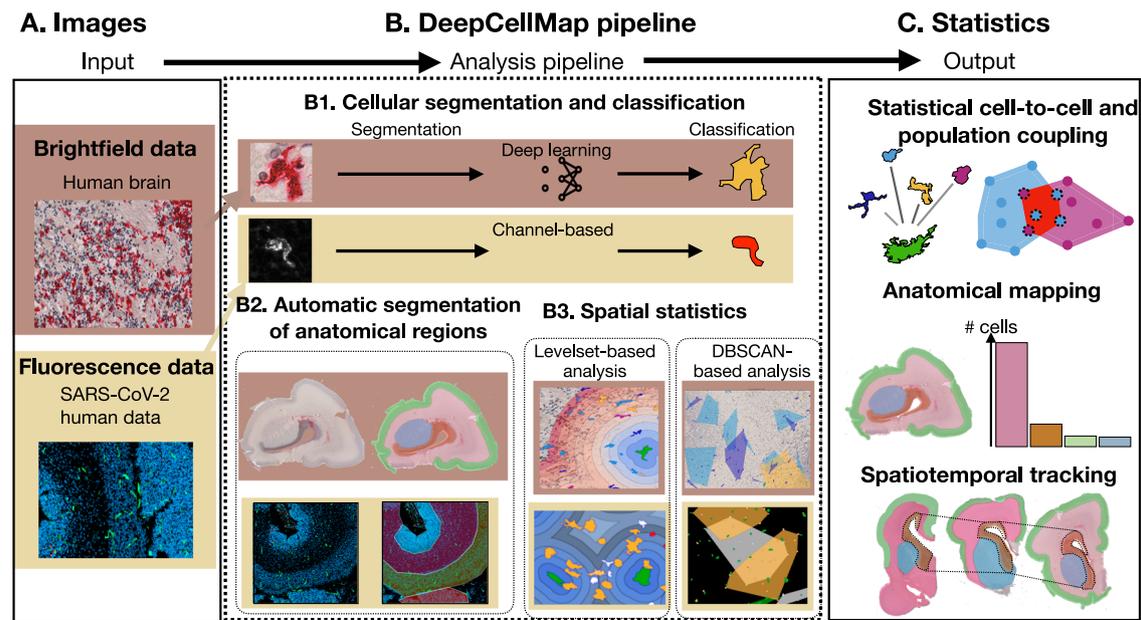
Fig. 1 | DeepCellMap principle. DeepCellMap is an open-source Python platform that segments, and classifies cells, and extracts spatial patterns within tissues. **A** Input can either be brightfield (e.g., immunohistochemistry data from a human developing brain) or multicolor fluorescent images (e.g., SARS-CoV-2 labeling). **B** First series of functions is dedicated to cell segmentation. The classification of cell morphologies in brightfield images uses a trained deep-learning network (U-Net), while the classification in fluorescent images is based on color channel separation (**B1**). To track the temporal changes of cellular distributions in different regions of the tissue during development, DeepCellMap embeds algorithms to automatically delineate specific tissue regions based on the global cell density (**B2**). From extracted cell types/morphologies and positions within the tissue, DeepCellMap computes spatial statistics to probe cell-to-cell and cell-to-region associations with levelset-based analysis, and the overlap between the different cell clusters with DBSCAN-based analysis (**B3**). **C** DeepCellMap provides detailed anatomical mapping of different cell types/shapes and enables the measurement of the accumulation of cells in different anatomical regions of the tissue. The spatial statistics allow the in-depth characterization of the association between different cell types (cell-to-cell and cluster overlap) and the evolution of these metrics across time.

Following cell segmentation, DeepCellMap classifies microglia into distinct morphological categories that likely reflect their types. In fluorescence imaging, cell types correspond to specific labels and color channels. For example, in our SARS-CoV-2 dataset, different cell types are associated with specific (combinations of) color channels. However, in brightfield imaging, identifying different cell types from the segmented masks is less straightforward.

Although there have been some attempts to automatically determine cell types using machine learning in well-defined imaging conditions[18], this process is still largely performed manually, particularly in developmental stages where diverse morphological types coexist within tissues[9,19]. DeepCellMap integrates a supervised deep learning classifier based on a U-Net architecture (see "Methods") to classify microglia into 5 distinct types (see Supplementary Figs. 7, 8, 9, and 10). The classification relies primarily on morphological features and information about the local cellular neighborhood in the tissue[34]. The classifier achieved an F1 score of 81% on the training dataset. More specifically, the proportion of well-classified cells in each class were proliferative: 85%, amoeboid: 85%, aggregated: 83%, phagocytic, 75%, and ramified 75%.

Delineation of brain regions. Automatic classification enables the collection of a large number of cells across different brain regions. The third step in the DeepCellMap pipeline involves the automated delineation of brain regions (Fig. 1B2). Distinct differences between cell densities exist between regions, and we developed an automated method based on the local density of detected cell nuclei (see "Methods"). Cell nuclei segmentation was performed using the well-established deep learning algorithm, CellPose[16], which facilitated accurate mapping of cell density across the imaged tissue. By applying automatic thresholding and morphological operations, we were able to successfully reconstruct four primary anatomical regions within the brain tissue: the striatum, neocortex, the ganglionic eminence, and the cortical boundary.

Mapping cell-to-cell spatial association. To map the spatial relationships between the cell morphologies/classes identified by the deep learning classifier (Fig. 1B3), as well as between cells and specific tissue regions, we developed a statistical framework that accounts for the local cell density and corrects for potential misclassifications from the previous deep-learning algorithm. Many existing methods for analyzing cell distribution in tissues focus on defining homogeneous cellular neighborhoods and extracting statistics, such as the accumulation of specific cell types within given regions[22,35,36]. However, few approaches offer a statistical characterization of spatial relationships across varying distances between cells[37]. Here, we build on the Statistical Object Distance Analysis (SODA) framework[38,39], which corrects for the non-significant spatial association of cell populations that may arise from random distributions, providing key metrics such as the percentage and mean distance between spatially associated cells. To examine the spatial distribution of one cell morphology B relative to another A, SODA employs a levelset method to map the spatial neighborhood of A cells, where the 0-level contour defines the boundary of A cells. The surrounding area is partitioned into subregions (ω_i, i = 1..N), and SODA adjusts for cell density variations to control for random spatial proximity (null hypothesis). This enables robust quantification of significant spatial associations between different cell types, or accumulation around a given tissue region, offering a deeper understanding of cellular organization and interactions.

Uncertainty in cell type classification can introduce errors that propagate into the spatial association analysis. For instance, in a scenario where type B is frequently misclassified as type C, any observed association between types A and C would, in fact, reflect the true association between types A and B. To mitigate these effects, DeepCellMap

incorporates classification uncertainty by weighing the cell counts in each level set region based on the output probabilities from the deep learning classifier (see "Methods"). Furthermore, we utilized the confusion matrix from the training dataset as an a priori correction, adjusting for potential misclassifications between cell types during the cell-to-cell spatial association analysis.

To validate the ability of DeepCellMap to accurately measure the level and distance of association between different cell types, while accounting for potential classification errors, we designed synthetic simulations (Supplementary Fig. 11). In these simulations, two cell types *B* and *C* were spatially associated with a third cell type *A* across a broad range of association parameters (level and distance). To model potential misclassification between cells, we created three scenarios: no confusion, intermediate confusion, and high confusion, where the classification error between *B* and *C* was set to 0%, 30%, and 45%, respectively (see "Methods"). The reconstruction errors, defined as the absolute differences between the association level and distance estimated by DeepCellMap (with and without correcting for misclassification) and the simulated parameters are presented in Supplementary Fig. 12 for the three confusion scenarios.

Overall, applying DeepCellMap without correction of classification errors leads to incorrect estimates of the association score (up to 40% error in case of intermediate confusion according to Supplementary Fig. 13A and 45% in case of high confusion (Supplementary Fig. 14C) and can lead to errors of up to 400 pixels for the evaluation of the association distance in both scenarios (Fig. Supplementary 13B, D and Supplementary Fig. 14B, D). We observed that the accuracy of DeepCellMap decreases as the standard deviation of the association distance increases, thus as associated cells are distributed across a larger number of level set regions, performance declines slightly. Nevertheless, while the algorithm tends to slightly underestimate the level of association, it maintains high accuracy in estimating distances, with an average error of less than 5% across the entire range of parameters tested. These results demonstrate that DeepCellMap reliably characterizes spatial associations between different cell populations, even in the presence of classification uncertainties and errors.

Cell cluster overlap. Level-set analysis is highly effective for probing cell-to-cell and cell-to-region spatial associations, but it has limitations in describing the relationships between the spatial territories occupied by different cell types within tissues. To address this issue, DeepCell-Map incorporates a semi-automatic DBSCAN-based algorithm, where one of the key parameters is automatically optimized (see "Methods"). To validate this function, we performed synthetic simulations where we varied the percentage of points organized in clusters and measured the performance of the optimized DBSCAN algorithm to retrieve the simulated proportion of clustered points (Supplementary Fig. 15). DeepCellMap was able to recover a large proportion of clustered points, with a slight underestimation probably due to the difficulty of taking into account points at the edge of Gaussian clusters.

Therefore, the present clustering analysis enables the robust detection of cell clusters, and computes the proportion of cells distributed in clusters *versus* isolated. The method also quantifies the degree of overlap over time and between different cell-type clusters. Combined with brain region delineation based on relative cell densities (as described above), this clustering analysis allows for spatiotemporal tracking of the colonization of various cell types during development (Fig. 1C).

Microglial colonization in the developing human brain
The colonization of the developing brain by different morphological types of microglia cells remains difficult to evaluate and could benefit from an automated identification of their spatial organization within different regions of the brain[5,7]. Using immunohistochemical labeling of microglial cells and brightfield imaging of post-mortem brain tissue at

different developmental stages ("Methods"), we used DeepCellMap to characterize distributions and spatial associations of different microglial phenotypes (morphological and functional) in different brain regions.

Microglial quantification in distinct anatomical regions. Using DeepCellMap, we automatically delineated and segmented the striatum, neocortex, cortical boundary, and ganglionic eminence (Fig. 2A1, A2) and extracted microglial counts by morphological class in each region using a segmentation pipeline based on Cellpose and Otsu methods (Supplementary Fig. 16 and Supplementary Fig. 17). We found (Fig. 2B) over three time points 17, 19 and 20 pcw that the total number of microglia was not continuously increasing over the covered time period. For example in the neocortex (pink labeled), the number first decreases from 10928 to 6926 and then increases to 14262. The number of microglia in the other regions was increasing.

We then quantified the proportion of the 5 classes of microglia in the same regions (Fig. 2C) using the deep learning classification approach integrated into DeepCellMap. We observed, from 17 to 20 pcw, an increase number of amoeboid microglia in the ganglionic eminence and a net increased proportion of phagocytic cells (see also Supplementary Fig. 18). The large proportion of amoeboid cells in the ganglionic eminence aligns with their role in migrating to colonize nearby structures such as the striatum. We also found that ramified cells are dominant in the striatum, neocortex, and cortical boundary, most likely because these regions are maturing at a faster rate than the ganglionic eminence, which remains mostly populated by amoeboid cells (Supplementary Fig. 19). Overall, DeepCellMap allows the classification of microglial classes according to morphology, the automated segmentation of anatomical areas according to cell density and the further characterization of the total and class-specific microglial cell counts in the different brain regions during development.

Spatial associations between microglial phenotypes. To study the spatial associations between microglial morphologies, we first applied the level set algorithm in DeepCellMap after choosing a region of interest and defining the states to be classified (Fig. 3A–C). Briefly, this method allows us to map the spatial neighborhood of a given cell class (morphological type) and extract the number of cells from another given class in level-set-defined regions. Importantly, the DeepCellMap implementation of level analysis corrects for the uncertainties in classification and the expected accumulation of cells in different level-sets for a random distribution (Fig. 3D).

First, we report here that all microglial morphologies are spatially associated with one another but at different levels. In particular, three microglial morphologies are most frequently associated with each other (phagocytic, aggregated, and proliferative) whilst ramified and amoeboid are less coupled with other cell morphologies. The level set analysis also allowed an estimation of the mean distance between associated single cells. For example, we found that the mean distances weighted by the level set (see "Methods") is 76 μm for ramified versus aggregated (Fig. 3E) in the striatum (Supplementary Figs. 20, 21, and 22). Interestingly the mean distance from all morphologies to the aggregated class was ~75.5 μm. The phagocytic class was the closest on average to the aggregated class with a mean distance of 68 μm. When we performed this analysis on all groups, we found that amoeboid, phagocytic, and ramified had mean distances to the other groups of 18, 26, and 29 μm respectively. This is in contrast with proliferative that were on average, at the largest distance of 143 μm from all other classes (Fig. 3F).

Second, we estimated the mean distances between the 5 microglial classes and the border of the brain region (Fig. 3G–I and Supplementary Fig. 23): we found that proliferative and aggregated microglia were the closest to the border with a mean distance of 202 μm, the others being at a distance of the border of ~238 μm on average. At this stage, we concluded that the levelset analysis reveals a
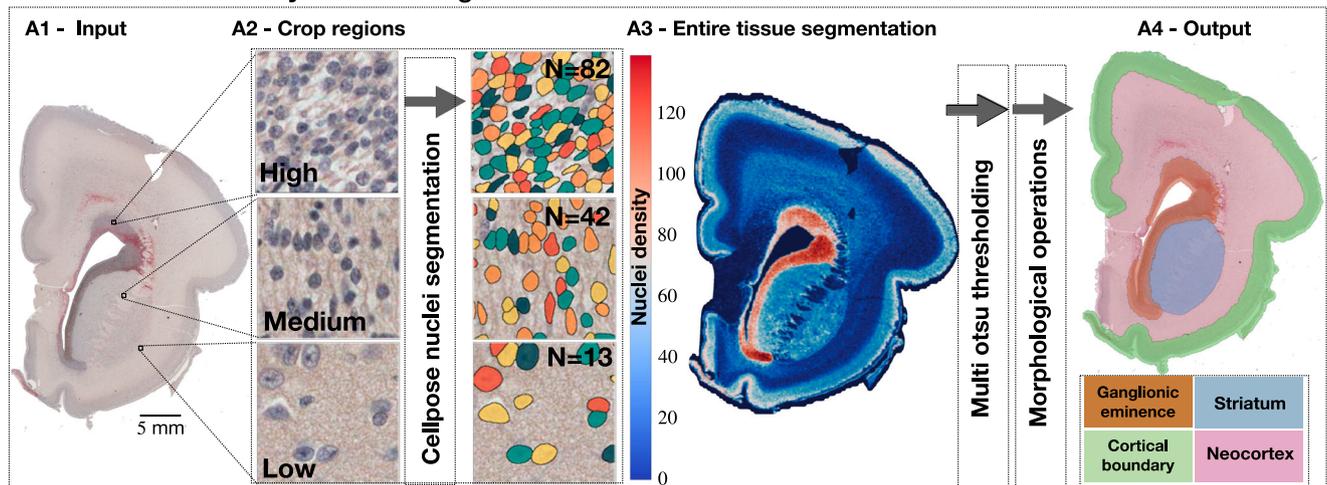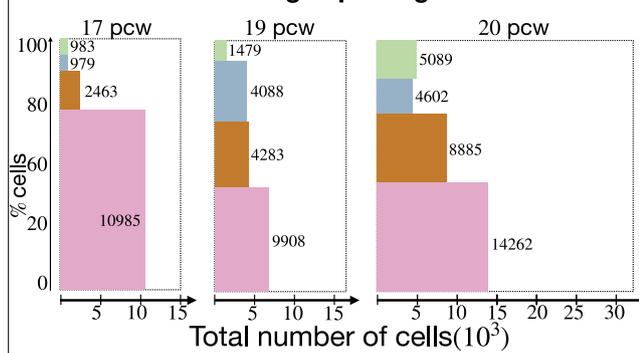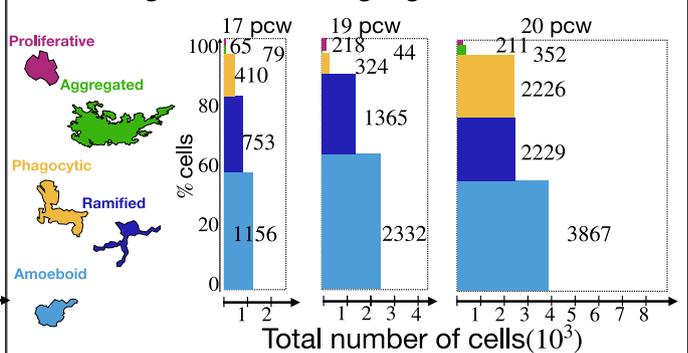
A. From nuclei density to tissue segmentation

B. Number of microglia per region

C. Microglial class in the ganglionic eminence

Fig. 2 | Tracking microglial colonization in different brain regions. A Detecting cell nuclei performed by CellPose segmentation algorithm[16]. **A1**- Image of a fetal tissue section at 17 pcw. **A2**- Cropped regions (256 × 256) showing three different densities (high, medium, and low), where we identified 82, 42, and 13 cells respectively. **A3** Nuclei density over the entire tissue (left color bar). This step is followed by the Otsu-thresholding and morphological operation pipeline (Supplementary Figs. 16 and 17). **A4** Automatic delineation of four brain regions based on nuclei density: striatum (blue), neocortex (pink), ganglionic eminence (orange), and cortical boundary (green). **B** Percentage and the number of microglia estimated from the deep learning algorithm for three pcws (17, 19, and 20) (*n* = 3). **C** Distribution of microglial morphologies in the ganglionic eminence. Source data are provided as a Source Data file.

certain variability of coupling, that departs from a uniform distribution, but this variability depends on the morphological class. Moreover, in the ganglionic eminence, the average coupling distance between amoeboid and phagocytic versus ramified (Fig. 3J1), was quite stable across time, with a mean distance of $29 \pm 4\,\mu$ and $61 \pm 2\,\mu m$ respectively. For other cell classes (proliferative and aggregated), the coupling frequency was less than 5% for 1 and 2 time points respectively, but when the frequency was higher, the coupling distances with ramified cells were of the same order of magnitude between the ganglionic eminence and the striatum for 17 pcw. However, the distance between the boundary of the ganglionic eminence and each population was changing over time, but the coupling frequency remained quite low, showing that this effect could be marginal (see frequency of coupling tables inside Fig. 3J2).

To conclude, our key findings here suggest that three microglial morphological classes are most frequently coupled with one another (phagocytic, aggregated and proliferative) whereas ramified and amoeboid are less coupled in general. Whilst the coupling distance remained stable for the one-time point between the striatum and the ganglionic eminence, it doubled when we considered the ganglionic eminence border versus the cortical tissue border. Importantly, on average, the cell-to-cell coupling distances were much smaller (< 100 μm) than the average cell-to-region border coupling distances (200 μm on average and could reach 500 μm). In the context of our

knowledge of the rodent literature, microglial proliferation may be associated with increased apoptosis which could explain the strong coupling between our phagocytic and proliferative classes[40]. Microglia are phagocytes, and it is precisely during these temporal windows that we expect an increase in cell death as the brain prunes extranumerary cells[5]. Dying cells influence the distribution of microglia[41], and here, we report the coupling association between proliferative and phagocytic and aggregated microglia in humans. We also observe an overall increase in phagocytosis genes as development progresses (Supplementary Fig. 6B–D and Supplementary Table 2), which mirrors our histology data. Aggregated and phagocytic microglia accumulate very closely to the border and a drastic decrease of association at 20 pcw is visible, possibly linked to distinct territories being colonized by the different types. Finally a decreased accumulation from the ganglionic eminence border may suggest that cells could be migrating from the ganglionic eminence towards other regions of the brain.

Identifying distinct territories occupied by microglia during brain colonization. To further assess whether microglial morphological types could be non-uniformly distributed, we developed and applied our semi-automatic DBSCAN clustering algorithm ("Methods"). We obtained an ensemble of clusters for each morphological state delimited by their convex hull, as illustrated with the ganglionic eminence (Fig. 4A, B).
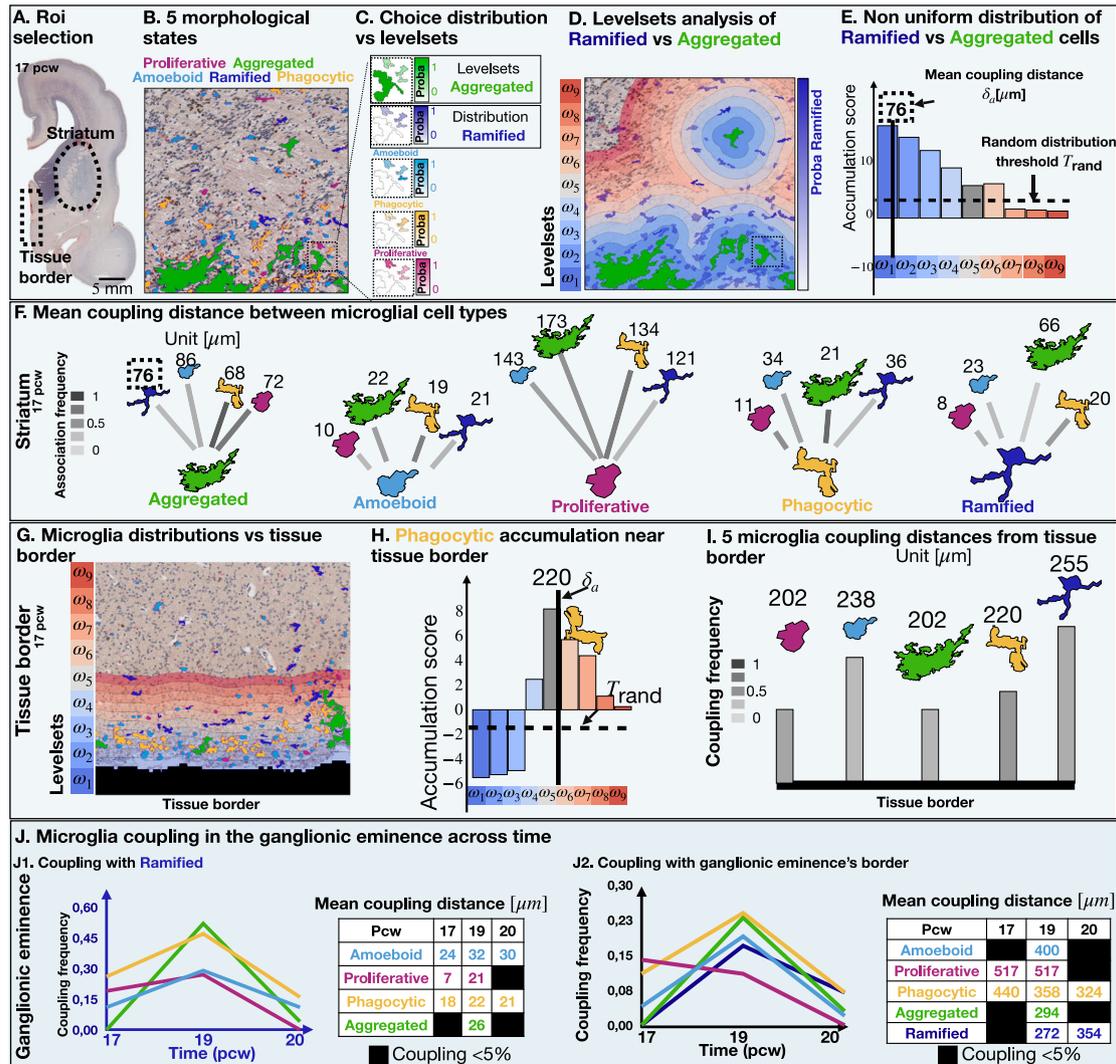
Fig. 3 | Microglial distributions according to levelset segmentation. A Choice of 2 regions: Striatum and tissue border in the fetal brain at 17 pcw. **B** Five microglial morphologies: proliferative, aggregated, phagocytic, amoeboid, and ramified, 2 of which are selected. **C** Choice of "aggregated" microglia as the center of levelset construction vs "ramified." **D** Ramified localization color-coded by probability (right blue scale) in the level set representation of the aggregated cells (green), obtained from a gradient distance. The distances are divided into 9 subregions ($\omega_1, ..., \omega_9$), color-coded (left scale). **E** Statistics of ramified cell accumulation in the levelsets. Accumulation-score is defined as the ratio $\frac{(K-\mu)}{\sigma}$, where K is the generalized Ripley function based on U-Net output probabilities ("Methods") and the expected mean μ and the standard deviation σ of ramified cells in each region ω_i. Statistical threshold $T_{rand} = \sqrt{2\log(n)}$ (horizontal dashed line), where n is the number of level sets: it is used to identify the ω_i regions where the microglial distribution deviates from the uniform one. Mean coupling distance $\delta a = 76\,\mu m$ between ramified and aggregated measures the mean distance between two populations, with weights accounting for the deviation ω_i in microglial distributions ("Methods"). **F** Coupling result is shown as trees, where the root (level sets) is a chosen microglial morphology, and the leaves (microglia distributed in the level set) are the 4 others, quantified by the mean coupling distance in μm (above) in the striatum. The distance lines are color-coded by the frequency of association (left scale bar). **G** Microglial cells in the level sets generated by the edge of the tissue or a region. **H** Distribution of phagocytic cells in the level sets ω_i, with the same threshold as above. **I** Coupling distances and association frequencies for each morphological type with respect to the edge of the tissue. **J** Spatiotemporal coupling of four microglia morphologies vs ramified (level set segmentation) in the ganglionic eminence (interactions where coupling frequency is below 5% are discarded) (**J1**), and coupling of 5 microglia shapes with respect to the edge (**J2**) across time. Source data are provided as a Source Data file.

We observed significant variability in the percentage of cells organized in clusters *versus* isolated between the different morphological types, with phagocytic and ramified cells being predominantly organized in clusters with 65% and 64% of cells found in clusters respectively (Fig. 4C, D) unlike amoeboid cells where the fraction of cells organized in clusters drops to 46% (see also Supplementary Fig. 24 for the other regions). We also investigated the changes over time in the fraction of clustered cells (Fig. 4E) for the 5 microglia types in the four identified anatomical regions (striatum, neocortex, cortical boundary, ganglionic eminence): we found quite homogeneous fractions of clustered cells in the ganglionic eminence (Fig. 4E), suggesting that specific classes tend to group in specific territories.

To further study the mixing of different microglial classes, we estimated the fraction of the clustered cells A in the intersecting area formed by clusters from A and an other morphological category B by the total area of these clusters (Fig. 4F). We then estimated in the ganglionic eminence the fraction of cells from one type in the clusters of the other types, as summarized in the matrix representation of Fig. 4G. Interestingly, we found that 44% of proliferative territories (not isolated cells) were mixed with aggregated territories (clusters). Also, there were no clustered phagocytic and aggregated cells in proliferative clusters, and no clustered aggregated in the ramified clusters. This suggests that only certain morphological types, when distributed in clusters, can mix but others do not, a situation that should be further explored.
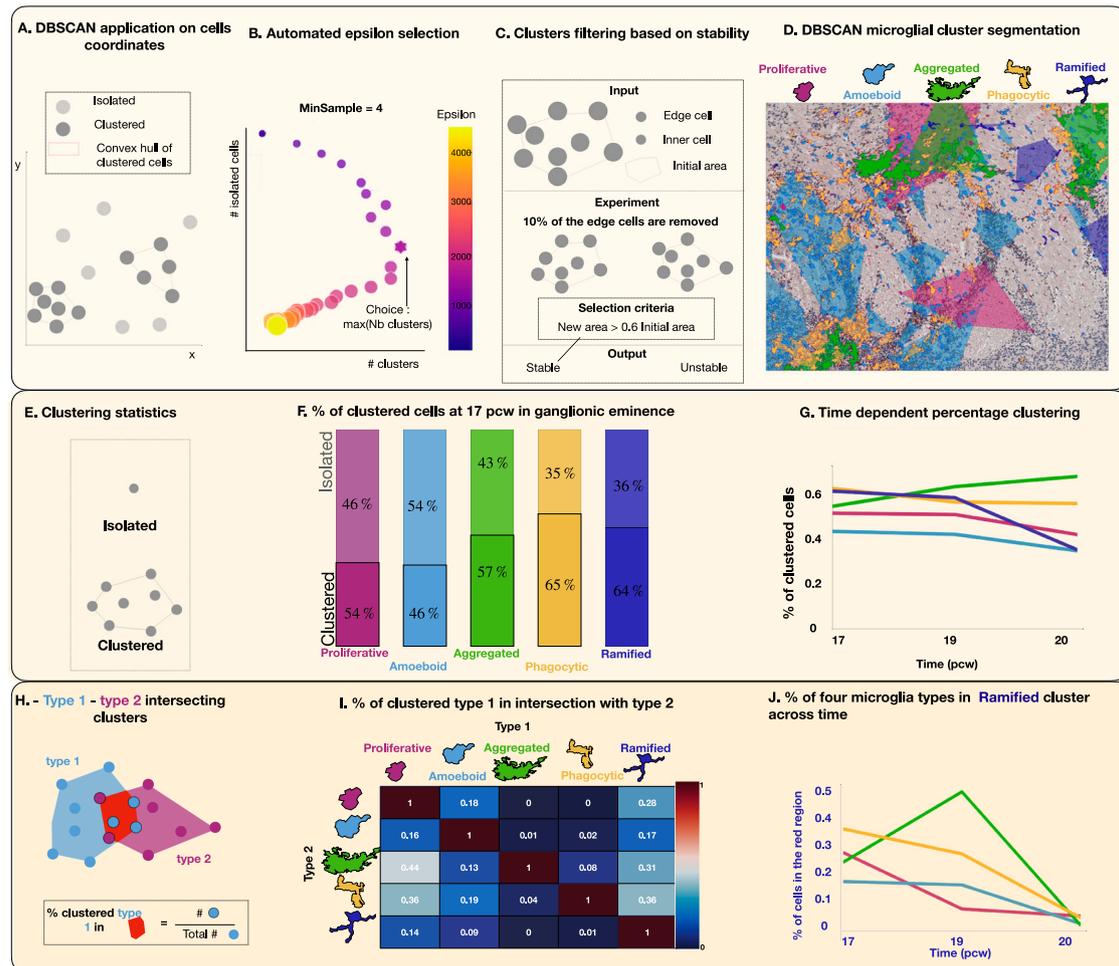
Fig. 4 | Overlapping of cell clusters in the ganglionic eminence. A DBSCAN application to the (x,y) coordinates of the cells' center of mass. **B** Graph of Min_{Sample} versus ϵ used to select the value of the radius ϵ that maximizes the number of clusters. **C** Schematic representation of the procedure to select stable clusters by removing 10% of the cell located on the edges of the convex hull while the area of the new convex hull remains > 60% of the initial surface. **D** Clustering of the five morphological states using DBSCAN algorithms, generating convex hulls around each cluster. Color code is the same as microglial morphological types: proliferative (pink), amoeboid (light blue), aggregated (green), phagocytic (yellow), and ramified (dark blue). **E** Scheme of isolated vs clustered cells. **F** Percentage of isolated vs clustered cells for the five populations at 17 pcw in the ganglionic eminence. **G** Time-dependent clustering percentage at 17,19, and 20 pcw. **H** Scheme of intersecting clusters from two different populations (type 1 vs type 2) and associated metric. **I** Cluster mixing rate matrix between the 5 microglial morphological states in the ganglionic eminence at 17 pcw. **J** % of four microglia types vs ramified cluster across time at 17,19 and 20 pcw. Source data are provided as a Source Data file.

Finally, we explored the time evolution of mixing between the different microglial types (Fig. 4H). This analysis allows us to follow in time how cell clusters are overlapping. We chose ramified cells as a reference and computed the fraction of the four other microglial types located inside ramified clusters. We found that from pcw 17 to 20, the percentage of all groups inside the cluster of ramified decays below 10% (this trend is similar for the striatum Supplementary Fig. 25, but not for the neocortex Supplementary Fig. 26). In the cortical boundary, this ratio is also around or less than 10% (Supplementary Fig. 27).

To summarize, in all 5 morphological classes and the 4 regions we analyzed across time, our findings suggest that there is no strong repulsion or attraction of clusters by other clusters. Furthermore, as development progresses, the overlap between the different classes is reduced so that each morphological group occupies a specific territory. This is consistent with brain development but not reported in humans. As the structures mature, we report here the switch to a more ramified microglia phenotype for example, and this is captured by our single-cell data too (Supplementary Fig. 6). From the mouse, we know that microglia are driven to occupy the space during embryonic development by proliferation but as the proliferative potential drops,

contact inhibition which refers to the physio-regulatory mechanism of arresting cell division when two cells come into physical contact, may provide a mechanism through which microglia occupy space with no overlap[9]. This is consistent with the distinct morphologies occupying specific niches in humans from our data. Furthermore, the underlying neurodevelopmental landscape may drive a specific phenotype where the most striking separation between clusters is observed in more mature structures (e.g., striatum) rather than transient ones (e.g., ganglionic eminence) where there is more overlap between clusters.

Microglial classes correlate positively with proportions obtained using scRNAseq analyses. We identified 10 clusters (Supplementary Fig. 6C) and annotated microglia and other cell types according to gene lists obtained from published datasets (Supplementary Table 2). Our purpose here was to demonstrate whether proportions obtained from DeepCellMap within similar pcws correlated with single-cell obtained proportions of the same classes of microglia. We, therefore, chose the three most functionally distinct classes - homeostatic, phagocytic, and proliferative. Amoeboid and aggregated microglia though morphologically distinct, may have dual functions - migratory or phagocytic/
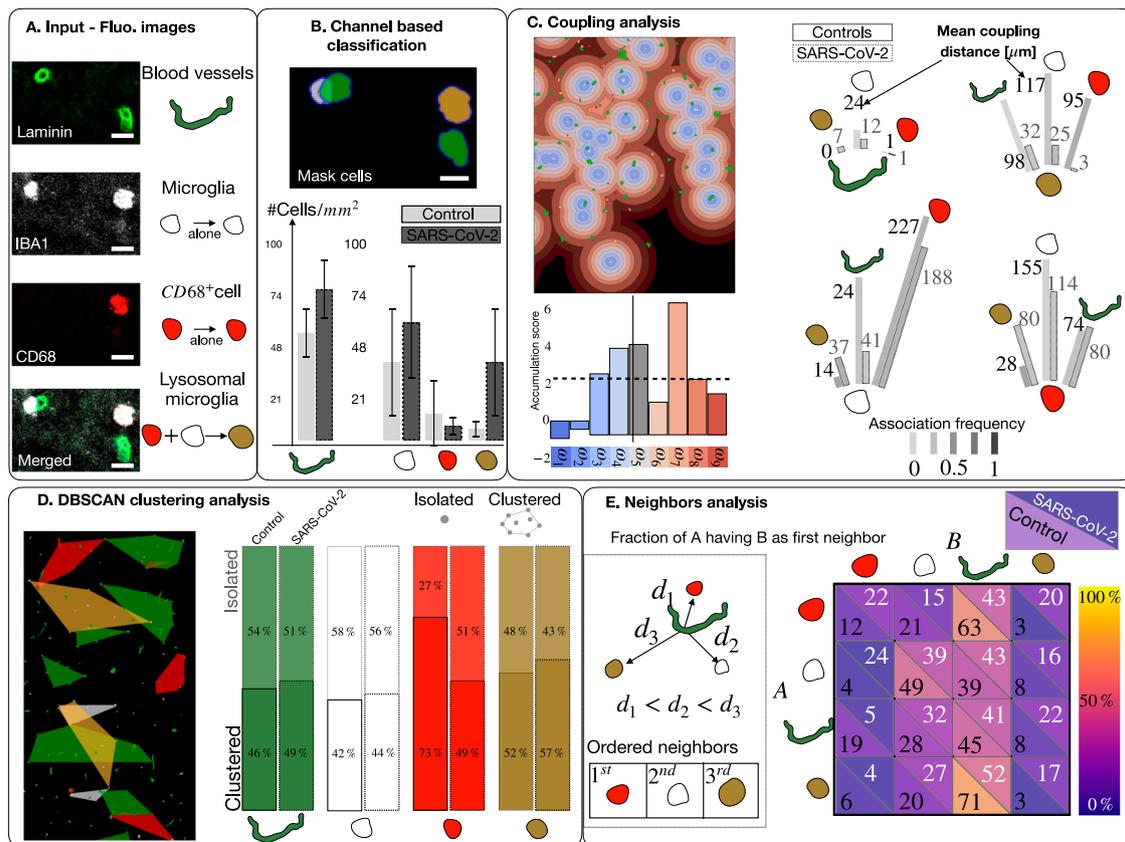
Fig. 5 | DeepCellMap applied to SARS-CoV-2 (COVID19) fetal human brain confocal data. A 3 channels : Laminin (blood vessels), IBA1 (microglia/macrophages), and CD68 (Lysosomal marker). White scale bar = 10 μm. **B** Cells are classified according to the channels from which they are segmented. We segmented four types of objects: blood vessels from channel Laminin (green), microglia from channel IBA1 (white), lysosomal cells from CD68 channel (red) and phagocytic microglia from the superposition of CD68 and IBA1 channels (n = 3) controls and

n = 4 SARS-CoV-2 exposed sample. Error bars correspond to standard errors (all values are available in Source Data). White scale bar = 10 μm. **C** Coupling frequency and mean distances. **D** Fraction of isolated vs clustered cells computed from the DBSCAN algorithm. **E** Fraction of (**A**) cells having (**B**) cells as first neighbors. Differences between control and SARS-CoV-2 are represented in the matrix. Source data are provided as a Source Data file.

activated. Albeit it being a restricted temporal window, we are able to show with Spearman's correlation that the proportions of microglia calculated by DeepCellMap in 3 classes (ramified/homeostatic, phagocytic, and proliferative) in the forebrain correlate positively with proportions of microglia calculated from our single-cell RNAseq data (homeostatic, phagocytic and proliferative) matched for 3 pcws weeks 11,12 and 14 (r = 0.64, p = 0.06) (Supplementary Fig. 6E). If we take the proliferative and ramified/homeostatic, the correlation is even stronger (r = 0.85). This shows a good positive correlation between DeepCellMap and single-cell-transcriptomic proportions in two independent datasets.

Robust association between microglia and blood vessels in response to SARS-CoV-2 in human fetal cortex

To show the versatility of DeepCellMap, we used our pipeline in fluorescence imaging of microglia in neocortical tissues of patients infected with SARS-COV-2, labeled using antibodies against SARS-CoV-2, microglia (IBA1), blood vessels (Laminin) and lysosomal cells (CD68) ("Methods"). Using image processing tools (Methods), we detected and classified four types of cells, to the channels from which they are segmented: blood vessels from channel laminin (green), microglia from channel IBA1 (white), lysosomal cells from CD68 channel (red) and phagocytic microglia from the superposition of CD68 and IBA1 channels (Fig. 5A).

In fluorescence microscopy, the morphological diversity of microglia is reduced from the classes we observed in brightfield during normal neurodevelopment. IBA1 cells were mostly amoeboid/round

microglial cells consistently across multiple regions including the neuroproliferative zones, the cortical plate, the subplate, and the marginal zone of the telencephalic wall (Fig. 5B). We extracted cell counts for each object identified above and normalized these to tissue size to obtain a 2D density in cells/mm2. We found that in SARS-CoV-2 samples, an increased blood vessel density, phagocytic microglia/macrophages, and resting microglia compared to controls which was consistent with figures obtained using manual counts in a larger dataset (n = 3 controls and n = 4 SARS-CoV-2 exposed samples), as shown in Fig. 5B[14]. Importantly, about 50% of microglia were IBA1 + CD68 + in SARS-CoV-2 samples, compared to 9% in controls (Fig. 5B).

To test whether any of the identified cell organizations departed from a uniform distribution, we then measured the spatial association between the different cells (microglia, lysosomes CD68 + cells, and lysosomal microglia) and blood vessels (Fig. 5C). We observed that the distance between spatially associated cells, and between cells and blood vessels was significantly reduced in SARS-CoV-2 samples compared with controls. We then quantified the number and overlap between the different cell clusters in SARS-CoV-2 (Fig. 5D). We measured an important decrease of CD68 + cells clustering in SARS-CoV-2 samples (49% of cells in clusters) compared with controls (73 %). We also observed a substantial overlap between lysosomal microglia (IBA1 + CD68 +) clusters and blood vessels. Nearest-neighbor analyses (see Methods) confirmed the tight association between lysosomal microglia and blood vessels, in both SARS-CoV-2 and controls (Fig. 5E).

Overall, the spatial statistics computed by DeepCellMap (spatial association, clusters' overlap and nearest-neighbor analysis) all point to a stronger association of lysosomal microglia, with the vasculature in SARS-CoV-2 samples, suggesting that these microglia could be sending or receiving signals which may alter vessel integrity which has yet to be probed further as a hypothesis. Overall, DeepCellMap is a versatile tool that can be extended to fluorescence confocal images to extract robust metrics about the spatial distribution of cells in control and disease conditions.

Discussion

To investigate the spatial distribution of cells in digital pathology and extract organizational patterns, we developed a computational pipeline, available as the open-source platform DeepCellMap. This platform integrates deep learning for cell classification with advanced spatial statistics. DeepCellMap corrects for possible confusion between cell morphotypes during automatic classification, and local cell densities during statistical parameter estimations, ensuring robust analysis. This pipeline was designed to map microglial organization during normal and pathological brain development but has the potential to be adapted to any cell type.

In the age of digital pathology, DeepCellMap allows the rapid extraction of information about core inflammatory processes to probe specific hypotheses. It can be adapted in clinical and scientific settings, identifying basic counts in different classes of cells, statistical associations between pairs, and cluster analyses which places it along the continuum from a basic diagnosis based on an inflammatory profile to a platform that allows the probing of robust patterns/characterization of cell populations for scientific studies. Because of its dual potential and our expanding it to fluorescent images, it can be widely used.

With DeepCellMap, we capture here the morphological diversity of microglia during human brain development and establish a thorough, and adaptable pipeline to group them into morphological classes. We know from the rodent that the exponential expansion phase of microglia during murine development is largely postnatal and more linear prenatally following the growth of the brain[9,41]. From our data, although restricted by the temporal windows we had investigated, we capture the decrease in proliferative microglia as development progresses. We also identify that whilst in some regions the numbers keep increasing (the striatum, the ganglionic eminence), in others (e.g., the neocortex), they undulate consistent with previous literature[5,9,42] suggesting that the trend of expansion and refinement of the population will vary by anatomical region.

In terms of the occupation of space by microglia, as development progresses, the overlap between the different classes is reduced so that each morphological group occupies a specific territory. This is consistent with brain development but not clearly reported in humans. As anatomical structures mature, we report here the switch to ramified microglia, for example and this is captured by our single-cell data as well. From the mouse, we know that microglia are driven to occupy the space during embryonic development by proliferation, but as the proliferative potential drops, contact inhibition which refers to the physio-regulatory mechanism of arresting cell division when two cells come into physical contact, may provide a mechanism through which microglia occupy space with no overlap[9]. This is consistent with the distinct morphologies occupying specific niches in humans from our data. Furthermore, the underlying neurodevelopmental landscape may drive a specific phenotype where the most striking separation between clusters is observed in more mature structures (e.g., striatum) rather than transient ones (e.g., ganglionic eminence) where there is more overlap between clusters.

The strong coupling association between morphotypes such as proliferative and phagocytic or aggregated is in line with the rodent literature that microglial proliferation may be associated with increased apoptosis[43,44]. Microglia are phagocytes, and it is precisely during these temporal windows that we expect an increase in cell death as the brain prunes extranumerary cells. Dying cells influence the distribution of microglia, and we report here the coupling association between proliferative phagocytic and aggregated microglia in humans. We also observe an overall increase in phagocytosis genes as development progresses (single-cell data), which mirrors our histology data Supplementary Fig. 6B–D and Supplementary Table 2). Overall, DeepCellMap provides new metrics that are informative regarding the overall organization of a very complex cell type during human development.

DeepCellMap revealed the association with blood vessels in SARS-CoV-2 samples, and this is particularly significant for two reasons: first, SARS-CoV-2 samples exhibited weakened blood vessels due to the loss of claudin-5 tight junctions, and second, these samples contained hemorrhages[14]. The increase in microglial numbers and their association with blood vessels may suggest that microglia may either be responding to or contributing to, changes in vascular integrity. This raises important questions about the role of microglia in areas of hemorrhage or in blood-brain barrier disruption in other diseases[45]. Specifically, are microglia aiding in the repair of weakened vessels, or are they contributing to the loss of claudin-5. This warrants further investigation.

In terms of applications, DeepCellMap applied to neurodegenerative disease tissues can be used to characterize concomitant pathology (alpha-synuclein, tau, TDP43)[46] which can lead to better quantification of associations between specific microglial classes and neuronal pathology. This can lead to the potential targeting of a specific microglial class or the class with the strongest association with the pathology for therapeutics. Furthermore, another layer may be added for those diseases that look pathologically similar but are clinically distinct, hence why a tool like DeepCellMap would be very valuable in characterizing the inflammatory profile to distinguish between diseases. From an imaging point-of-view, 3D reconstructions of microglia in thicker tissue sections would provide valuable insights into the migratory phenotype, which remains particularly challenging to define in humans. This approach would allow us to better understand the role of migration in microglial colonization during development and how this process may be altered by SARS-CoV-2 infection, further enhancing the capabilities of DeepCellMap.

By computing a wide range of statistics on the spatial relationships between different cell states or types, it becomes possible to map more complex cellular assemblies involving multiple cell types, similar to approaches used for molecular assemblies in super-resolution fluorescence microscopy[39,47]. The development of such multivariate spatial analysis is especially relevant given the rise of spatial transcriptomics and multiplex imaging techniques[48], such as multiplex immunohistochemistry[49], imaging mass cytometry[50], and co-detection by indexing[22], which can identify dozens of cell types in situ. These advanced imaging and mapping techniques will ultimately enable the deconstruction of complex spatial networks of cell types within tissues, providing robust and precise predictive models of clinical outcomes or, more broadly, revealing the biological principles that govern tissue organization.

Methods

Ethics statement and histological slides We labeled and selected high-resolution histological slides across human development from the 10th pcw-term for this study. Immunohistochemistry steps were consistent with the methods below for tissue processing for a total of 31 cases (12F;14M;5n/k). All demographics are summarized in Supplementary Table 2. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethics committees of the School of Medicine, University of Zagreb (protocol number 251-59-10106-23-111/158) and the Oxford Brain Bank (Rec approval: 23/sc/0241, South Central Oxford C). Part of the human fetal material for the

SARS-CoV-2 samples was provided by the Joint MRC/Wellcome Trust grant #099175/Z/12/Z Human Developmental Biology Resource. Ethical approval was granted by the Human Developmental Biology Resource (HDBR) jointly funded by the MRC and Wellcome Trust (Rec number (Newcastle): 23/NE/0135, Newcastle and North Tyneside ethics committee and Rec number (London): 23/LO/0312, Fulham ethics committee). Fetal scans, including the SARS-CoV-2 dataset summarized below, are available upon request. Written informed consent was obtained from all mothers for the use of these samples. Exclusion criteria were congenital abnormalities, genetic disorders, brain trauma, periventricular leukomalacia, and hypoxic ischaemic encephalopathy, except for the SARS-CoV-2 samples that had haemorrhages described elsewhere ([14]. Additional exclusion criteria were infection and brain trauma.

Anatomical areas
We focused on the forebrain along the frontal axis, which includes the cortex, the striatum, border regions, and the ganglionic eminence. The striatum's development begins at the basal telencephalon as early as the 7th pcw and spans the entire developmental period until 25–28 pcw[51,52]. The ganglionic eminence by that stage resolves and after 30 pcw, the corpus striatum seems to have reached its mature form. Nissl and PAS-AB labelings were used for the delineation of the anatomical boundaries of these structures (Fig. 1 and Supplementary Fig. 1A).

Immunohistochemistry. Paraffin-embedded blocks were cut into thin sections of 10 μm on a microtome for immunohistochemistry. Brightfield immunohistochemistry experiments were performed using antibodies against microglia with the following dilutions: rabbit anti-IBA1 (019-19741, Wako) or mouse anti-IBA1 (ab283319, Abcam); mouse anti-Ki67 (GA62661, Dako Omnis, Agilent), rabbit TMEM119 at 1:500 (ab185333, Abcam, mouse PG-M1 at 1:400 (GA61361-2, Dako Omnis, Agilent), mouse P2RY12 at 1:1000 (ab180366, Abcam) and rabbit SOX2 at 1:1000 (sc365823, Santa Cruz). The first step was deparaffinization of formalin-fixed paraffin-embedded sections in xylol solution and rehydration in descending concentrations of diluted ethanol (100%, 96%, 70%). Antigen retrieval was done by heat-induced epitope opening using citric acid buffer (pH = 6.2). Thereafter, sections were pre-treated with methanol and hydrogen peroxide to block endogenous peroxidase and phosphatase activity. Sections were blocked with a solution of 5% Bovine serum albumin + Tween20 (0.1%) + normal horse serum (5%) in 1X PBS and then incubated with primary antibodies overnight. The next day, secondary antibodies were applied using either the Immunopress duet kit (MP7724, Vector labs, UK) with anti-mouse epitopes visualized with DAB chromogen in brown and anti-rabbit epitopes visualized with alkaline phosphatase in magenta or the Envision kit (mouse/rabbit) (K500711-2, Agilent, UK). Sections were counterstained with haematoxylin or methyl green and coverslipped with a permanent mounting medium before imaging.

Brightfield slide-scanning. Imaging was done using high-resolution histological slide scanners: the Aperio Imagescope (Oxford, UK) and Hamamatsu Nanozoomer (Zagreb, Croatia) 2.0 RS with a 40x objective (numerical aperture of the lens = 0.75) at 0.45 μm × 0.45 μm pixel resolution.

Microglial morphologies selected. We sampled microglial morphologies based on published references during human development[5,7,28–33,53]. In brief, microglial morphologies described here were proliferative, amoeboid, aggregated, phagocytic, and ramified (Supplementary Fig. 1A).

For the coupling of morphology with function, we refer to the representative examples provided in Supplementary Fig. 6A. The dataset consists of a set of 24 histological images of human fetuses at various stages of brain development. Images were processed coronally

through the frontal axis of the brain from 10 post-conceptional weeks until term. All sections were histochemically labeled with haematoxylin and eosin (H&E) according to standard methods and assessed by a neuropathologist for histology.

Fluorescence data collection
Human tissues. We extended our brightfield pipeline to allow the processing of human fetal images using fluorescence microscopy. Demographics of the cases are provided in Supplementary Table 1. Ethical approval was granted by the Human Developmental Biology Resource (HDBR) jointly funded by the MRC and Wellcome Trust (Rec number (Newcastle): 23/NE/0135, Newcastle and North Tyneside ethics committee and Rec number (London): 23/LO/0312, Fullham ethics committee).

The HDBR provided fresh tissue from fetuses aged 9–21 post-conception weeks (pcw): we selected here from 26 haemorrhagic samples, 7 samples with elective terminations (3 with no histological abnormality recorded and 4 with cortical haemorrhages whose mothers were positive for SARS-CoV-2) for the purposes of demonstrating how DeepCellMap can be applied to these tissues. The processing of these samples has been specified elsewhere[1]. Briefly, all tissues were fixed for at least 24 h at 4 °C in 4% (wt/vol) paraformaldehyde (PFA) in 120 mM phosphate buffer (pH 7.4). Brains were then sucrose-treated (15 and 30% sucrose solution sequentially for 24 h each), OCT-embedded, and then 20 μm thick sections were cut using a cryostat.

Immunofluorescence. Sections were then stained in a solution containing 10% BSA and 0.1% Triton, using the primary antibodies CD68 (1:100 DAKO M0814), IBA-1 (1:100 Abcam ab5076), pan-laminin (1:100 Sigma L9393), rabbit IgG isotype control (1:50 Abcam ab172730 [EPR25A]), SARS-CoV-2 spike protein (1:100-250 Genentech GTX632604 [1A9]) and SARS-CoV-2 nucleocapsid protein (1:50-150 Sino Biological 40143-R001). Life technologies secondary antibodies, used at 1:1000, were donkey-anti-goat Alexa fluor 488 (A11055), anti-mouse Alexa fluor 568 (A10037) and 555 (A31570), anti-rabbit Alexa fluor 647 (A31573) and 488 (A21206), and goat anti-rat 555 (A21434). Sections were all stained with DAPI (Sigma D9542) and mounted in Mowiol (Merck Biosciences).

Fluorescence data and confocal imaging. Highly-resolved confocal imaging was subsequently performed using a Zeiss LSM 800 inverted microscope and a Zeiss Plan-Apochromat 20 × 0.8 objective, or a Zeiss AxioScan slide scanner and a Zeiss Plan-Apochromat 20 × 0.8 M27 objective at 0.312 μm/pixel as resolution. More details have been specified elsewhere[14]. We illustrate our image processing pipeline in Fig. 1.

Single-cell RNA-seq analysis of an independent dataset
We analyzed a total of 6584 microglial cell transcriptomes from the forebrain, midbrain, and hindbrain and focused our analyses on the forebrain, diencephalon, and telencephalon regions for the purposes of this work. These data were derived from[54]. The temporal windows considered were 10–15 pcw or the late first/early second trimester matching part of our histological temporal windows. Normalization and dimensionality reduction were done using Sctransform (version 0.4.0) and the RunPCA function of the Seurat R package (version 4.1.4) with default parameters. To harmonize the data, we used the harmony R package (version 1.0.3) to regress out the effect of each sample with max_{iter} set to 20 and the first 30 principle components. The Uniform Manifold Approximation and Projection (UMAP) plots were made using the RunUMAP function of the Seurat R package with dimensions set to the first 20 and using the harmony-derived dimensionality reduction. The number of dimensions for UMAP plots were defined using the Elbowplot function implemented in the Seurat R package. The scRNAseq clustering was done using

FindNeighbors (k.param = 15 and dims = 1:20) and FindClusters (resolution = 0.5, algorithm = 2) functions of Seurat R package. The resolution was defined using the *Clustree* R package (version 0.5.0). Annotation of clusters was done based on canonical and functional markers from[55–60], see Supplementary Table 2. We then calculated the proportions of 3 microglial classes extracted from our single-cell analysis and correlated these with the proportions of the same classes obtained using DeepCellMap matched for pcw(s) 10 to 14 (see Supplementary Fig. 6B–D).

Image processing

Pre-processing. Images were manipulated at different scales. First, tissue was extracted using Otsu thresholding and morphological operations on the downscaled image (factor 1/32 for brightfield and 1 for fluorescence data). A subdivision of the original images into tiles of size 1024 × 1024 pixels characterized by their row and column numbers allowed manipulation of the regions of interest at high resolution (Supplementary Fig. 3). Thus, each region of interest is characterized by its coordinates (top left tile and bottom right) in the image grid. Cells are identified by the quadruplet (*row*, *col*, *x*, *y*) describing the tile (*row*, *col*) to which their center of mass belongs and their coordinates (*x*, *y*) in that tile.

Cell detection and segmentation. Given the arbitrary division of the image into tiles and the possibility that cells may be present on several tiles, the following procedure is applied to ensure that all cells are taken into account and not counted more than once. During the detection procedure on a tile, the 8 neighboring tiles are considered, and the segmentation algorithm is applied to the resulting 3 × 3 tile image. The centers of mass of detected cells with part of their body in the central tile are calculated; if this center of mass is in the central tile, then the cell belongs to this tile. The segmentation algorithm proceeds in several steps, applied to each tile whose tissue percentage exceeds 5% (Supplementary Fig. 4A). First, *Image binarization* is performed with the Otsu adaptive thresholding from the eosin color-deconvolution of the RGB image[61]. The algorithm evaluates the histogram of the eosin image and finds two critical thresholds, allowing it to separate three regions. Pixels belonging to the third region are assigned a value of 1, while the others are set to 0 (Supplementary Fig. 4B1). Small isolated fragments of the binary mask are removed by using a disk of radius $r = 1$. (Supplementary Fig. 4B2). Then, a *morphological dilation* with a disk of radius $r = 3$ is applied to the binary mask to connect the different binary masks belonging to the same cells (Supplementary Fig. 4B3). *Holes in the cell mask are filled* by a background reconstruction (Supplementary Fig. 4B4), and finally, a *size filter* is applied to cell masks, and cells with a size smaller than a threshold $T = 700$pixels) are filtered out (Supplementary Fig. 4B5).

During this segmentation procedure, two parameters can be adjusted by the user: the radius r of the dilation disk (between 2 and 4), and the maximum size of the cells, which depends mainly on the resolution of the image.

The output is a binary mask of segmented cells over the background. Five parameters are associated with each segmented cell: row and column of the corresponding tile, coordinates (*x*, *y*) of the cell center of mass, and cell size (sum of mask pixels equal to 1). For brightfield images, the result of the DL classification into the different types of microglial cells is also listed.

Validation of cell detection. To validate how DeepCellMap detects microglial cells in brightfield imaging, we randomly selected $N = 20$ tiles of size 3000 × 3000 from the dataset. For each tile N_i, a certified expert counted the number n_i^e of microglial cells. This number was then compared to the number n_i^d of cells automatically detected by DeepCellMap (Supplementary Fig. 5. We computed the overall error by

the following formula:

$$e = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{n_i^e - n_i^d}{n_i^e} \right|, \qquad (1)$$

and obtained an overall error $e_{mean} = 15 \le$ with std error $e_{std} = 2.4 \le$ over $N = 20$ annotated tiles.

Since the detection algorithm is sensitive to size filtering of detected cells, the 'Laboratory cell segmentation' notebook has been created to calibrate this critical value as best as possible according to the dataset.

Deep learning classification of microglial cells

To ensure that the training database of annotations in each class covers a maximum of the inter-class spectrum, cells are randomly selected from images at different times: further several tiles are also randomly selected, cells are detected on a tile, and a fraction of these cells is added to a database of unlabeled cells. Cells are patches (size 256^2) whose center is the cell's center of mass.

An ergonomic annotation tool allowed the expert to label each cell belonging to one of the five morphological states. This made possible to build up a base of annotations in each class. We illustrated in Supplementary Fig. 28 this dataset using the ramified microglia cell type.

The RGB images of the cells and the labels assigned to the cells were given as input to the UMAP dimension reduction algorithm in order to visualize the inter-class heterogeneity of the cells in the training database and to manually correct annotation errors (Supplementary Fig. 8). After the training set has been built up, the masks contained only one cell in the center, but other cells may be present in the periphery. These cells were added to the mask with an "other" label (Supplementary Fig. 9) to prevent them from being considered as background by the model. RGB images and completed masks were then shuffled and split into training (75%), validation (15%), and test (15%) sets. The test set was used for the evaluation of the final model.

During training, images belonging to the training set were submitted to image augmentation procedures (brightness changes, flips, crops, and rotations) and were applied to each image to expand data diversity and make the model more robust[62].

A convolutional neural network (CNN) based on U-Net architecture was selected for cell classification. The network consisted of a contracting and an expansive path. The contracting path consists of three applications of the following steps : two 3 × 3 convolutions (with zero padding), each followed by a rectified linear unit (ReLU), and a 2 × 2 max pooling operation with stride 2 for downsampling. For the fourth layer of the contracting path, a dropout of rate 0.5 was added between the 2 convolutions and the max pooling operation to constrain the fully connected layers and to reduce overfitting. Then two 3 × 3 convolutions (same configuration with ReLu activation and zero padding) are applied before a dropout (rate 0.5) that ends the contracting path. Every step in the expansive path consists of a 2 × 2 up-sampling of the feature map followed by a 2 × 2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3 × 3 convolutions, each followed by a ReLU activation. At the final layer, a 1 × 1 convolution is used to map each 64-component feature vector to the classes we identified: background, proliferative, amoeboid, aggregated, phagocytic, ramified, and detected.

"Adam" optimizer was used for optimization, the initial learning rate was set to 0.003, batch size was set to 3. The CNN was trained on an off-the-shelf NVIDIA GeForce GTX 1080 with 8 GB GPU memory, for 40 epochs, training time took about 10 h.

For each pixel of the image, the output of the CNN algorithm is the probability that this pixel belongs to one of the 5 pre-defined classes of

microglial cells (Supplementary Fig. 10B3). The overall probabilities of segmented cells were obtained by averaging the probabilities of the mask's pixels. There are some aggregated cells (< 1%) whose cell body was larger than the patch, in this case, the mask probabilities of the cell in the patch were assigned to the full mask of the cell. Finally, microglial morphology was assigned by selecting the label for which the maximum probability was achieved.

Automatic delineation of tissue regions

To identify the components of the tissue with similar nuclei densities, we first segmented automatically cells' nuclei. In fluorescence microscopy, we used the color channel corresponding to nuclei labeling. In brightfield microscopy, the identification of nuclei is less obvious, and we tuned the CellPose algorithm[16]) to segment the nuclei contours in different types of images (Fig. 2A). The output of the algorithm is the number of nuclei present in each image patch of size 256×256, allowing to reconstruct the heatmap of the underlying tissue density. We focused on 4 regions: striatum, neocortex, ganglionic eminence, and the cortical boundary (Supplementary Figs. 16 and 17).

For each tissue, the histogram of nuclei densities was then separated into 3 or 4 regions using the Otsu-multi-thresholding algorithm, which optimizes the choice of thresholds. The different tissue classes obtained with Otsu thresholding then underwent several elementary morphological operations to distinguish each of the physiological regions identified in the images: striatum, neocortex, ganglionic eminence, and the cortical boundary (Supplementary Fig. 29). The cortical boundary and the neocortex (which cover regions several hundred thousand pixels wide in the images), were subdivided into 4 sub-regions in which the calculations were carried out and recombined by weighting the results as described in Supplementary Fig. 29.

Level-set analysis of cell-to-cell spatial association

To measure the spatial association between different populations of cells (or cells to regions) and account for the uncertainties associated with the deep-learning classification of cell types, we implemented a generalized version of the statistical approach developed in ref. 38.

The method can be described as follows: we first specify K cell types $\mathbf{A} = \{A_1, \ldots, A_k, \ldots A_K\}$. During the deep learning classification, each cell type A_k ($1 \le k \le K$) is classified as type A_j with probability p_{jk}. We highlight that these classification probabilities can be estimated from the results of the deep-learning classification on the training dataset.

To compute the potential accumulation of A_m cells around A_l cells, we first select the cells with a maximum classification probability in the type A_l and map the domain of interest Ω (typically a predefined region of interest within the tissue slide) around the population A_l with a series of levelset regions $\bigcup_{1 \le i \le n} \omega_i$, with $\omega_i = \{\mathbf{x} \in \Omega | r_i \le |\mathbf{x} - \partial A_l| < r_{i+1}\}$ which is the region of Ω that contains points $\mathbf{x}$ at a distance comprised between r_i and r_{i+1} ($r_1 = 0 < r_2 < \ldots < r_n$) from the contour ∂A_l of A_l cells.

To measure the spatial association of type A_m of cells to type A_l, we use the ensemble of cell coupling (center of mass typically) $\{u_j\}_{1 \le j \le N}$, with $N = \sum_{k=1}^{K} |A_k|$ the total number of cells, and the probabilities $\{0 < \hat{p}_j^m \le 1\}$ of being classified as a A_m type, and measure the total number of A_m points within the levelset ω_i according to the estimator:

$$\hat{\eta}_{\omega_i}^l(A_m) = \sum_{j=1}^{N} \hat{p}_j^m \chi_i(u_j), \tag{2}$$

where the indicator function $\chi_i(u_j) = 1$ if $u_j \in \omega_i$, and 0 otherwise.

Correcting for cell misclassification.

Automatic cell classification leads to potential errors, and a correction of the statistical analysis of cell-to-cell association is thus needed to account for misclassification: indeed, in the worst case scenario where cells of type A_m would be

misclassified as type A_n, the association of A_m cells to A_l cells would actually reflects the association of A_n cells to A_l cells.

To derive the correction term, we consider the $K \times K$ confusion probability matrix $\mathbf{P}$, with elements $(p_{nm})_{1 \le n, m \le K}$ the probabilities for an A_m cell to be classified as an A_n cell Since each A_m cell within the level set ω_i has a probability p_{nm} to be classified as an A_n cell, the measured accumulation $\hat{\eta}_{\omega_i}^l(A_m)$ of A^m cells within ω_i around A_l cells is equal to

$$\hat{\eta}_{\omega_i}^l(A_m) = \sum_{k=1}^{K} p_{mk} \eta_{\omega_i}^l(A_k). \tag{3}$$

with $\eta_{\omega_i}^l(A_k)$ the *ground-truth* accumulation of A_k cells in ω_i around A_l cells. Note that we have $\eta^l(\omega_i(A_l)) = 0$ because we are estimating the association to A^l cells.

If $\hat{\mathbf{H}}_{\omega_i}^l = [\hat{\eta}_{\omega_i}^l(A_k)]_{1 \le k \le K}$ and $\mathbf{H}_{\omega_i}^l = [\eta_{\omega_i}^l(A_k)]_{1 \le k \le K}$ are respectively the vectors of measured and *ground-truth* number of cells within levelset ω_i around A_l cells, we can rewrite the equation in a matrix form as:

$$\hat{\mathbf{H}}_{\omega_i}^l = \mathbf{P} \mathbf{H}_{\omega_i}^l. \tag{4}$$

and we can invert the *ground truth* which can be estimated from the measured accumulation of cells based on the relation

$$\mathbf{H}_{\omega_i}^l = \mathbf{P}^{-1} \hat{\mathbf{H}}_{\omega_i}^l. \tag{5}$$

Based on the previous estimation of the accumulation of the different cell types in levelsets around a given population A_l of cells, we aimed at determining the cell types that significantly accumulate in each level set. Such estimation will allow us to characterize the spatial association between the different types of cells.

To assess whether or not the accumulation of A_m cells in levelsets ω_i, $1 \le i \le n$ around A_l cells is significant, we followed the methodology given in[63] and computed the n-dimensional Ripley vector

$$\mathbf{K}^l(A_m) = \left[K_{\omega_1}^l(A_m), \ldots, K_{\omega_n}^l(A_m) \right]'. \tag{6}$$

with

$$K_{\omega_i}^l(A_m) = \frac{|\Omega|}{|A_m|} \eta_{\omega_i}^l(A_m). \tag{7}$$

that we rewrite

$$K_{\omega_i}^l(A_m) = \frac{|\Omega|}{|A_m|} \sum_{j=1}^{N} p_j^m \chi_i(u_j), \tag{8}$$

where p_j^m is the probability that cell position u_j belongs to the A_m population of cells. It can be computed from the classification probabilities $\{\hat{p}_j^k\}_{1 \le k \le K}$ and the confusion matrix $\mathbf{P}$:

$$p_j^m = \sum_{k=1}^{K} p_{km} \hat{p}_j^k. \tag{9}$$

The estimated total number of A_k cells is given by

$$|A_m| = \sum_{j=1}^{N} p_j^m. \tag{10}$$

When cells A_m are coupled to cells A_l, we expect a significant accumulation of the A_m cells' positions in the neighborhood of A_l, *i.e.* in a subset of the levelset regions ω_i for $1 \le i \le n$. To rule out a fortuitous accumulation of B cells due to chance, it is necessary to characterize the distribution of $\mathbf{K}^l(A_m)$ under the null hypothesis of complete spatial randomness where the A_m cell coordinates would be randomly and

uniformly distributed over Ω according to the Homogeneous Poisson law. Under the null hypothesis of A_m randomness, each function $K^l_{\omega_i}(A_m)$ is normally distributed[38]

$$K^l_{\omega_i}(A_m) =^d \mathcal{N}\left(\mu^l_i, \sigma^l_i\right). \tag{11}$$

To compute the mean μ^l_i and standard deviation σ^l_i, we highlight that, if A_m cells are randomly distributed, $\chi_i(u)$ is a Bernoulli variable with mean

$$\mu^l_i = \frac{|\omega_i|}{|\Omega|} \tag{12}$$

and variance

$$\left(\sigma^l_i\right)^2 = \frac{|\omega_i|}{|\Omega|}\left(1 - \frac{|\omega_i|}{|\Omega|}\right). \tag{13}$$

Therefore, we obtained that the mean is given

$$\mu^l_i = \frac{|\Omega|}{|A_m|}\sum_{j=1}^{N} p^m_j \mathbb{E}\left[\chi_i(u_j)\right] = |\omega_i| \tag{14}$$

and the variance is given by

$$\left(\sigma^l_i\right)^2 = \frac{|\Omega|^2}{|A_m|^2}\sum_{j=1}^{N}(p^m_j)^2 \text{Var}\left[\chi_i(u_j)\right] = \frac{\sum_{j=1}^{N}(p^m_j)^2}{|A_m|^2}|\omega_i|(|\Omega| - |\omega_i|) \tag{15}$$

Under the null hypothesis, the vector $\mathbf{K}^l(A_m)$ is Gaussian:

$$\mathbf{K}^l(A_m) \overset{d}{=} \mathcal{N}(\mathbf{M^l}, \mathbf{S^l}), \tag{16}$$

with the mean $\mathbf{M^l} = \left[\cdots \mu^l_i \cdots\right]'$, and the covariance matrix is

$$\Sigma = \text{diag}\left[\cdots \left(\sigma^l_i\right)^2 \cdots\right]. \tag{17}$$

The accumulation of cells within the levelset region ω_i is statistically significant when the mean exceeds a threshold

$$K^l_{\omega_i}(A_m) > \mu^l_i + \tau(n)\sigma^l_i, \tag{18}$$

where $\tau(n) = \sqrt{2\log n}$ is the universal statistical threshold used for determining the relevant signal in a vector containing Gaussian noise[64].

From previous statistical thresholding (eq. (18)), we can now estimate the subset of level set regions $\{\omega_{i_r}\}$ where the observed accumulation of A_m cells is statistically significant.

To quantitatively describe the association of the set of cells A_m to the set A_l, we computed the number $\{\eta^l_a(A_m)\}_{\omega_i}$ of spatially associated A_m cells inside each level set ω_i:

$$\left\{\eta^l_a(A_m)\right\}_{\omega_i} = \frac{|A_m|}{|\Omega|}\mathbf{1}\left\{K^l_{\omega_i}(A_m) > \mu^l_i + \tau(n)\sigma^l_i\right\}\left(K^l_{\omega_i}(A_m) - \mu^l_i\right), \tag{19}$$

which corresponds to the significant overcount of A_m cells inside ω_i above the expected number of randomly distributed cells. We then defined a probabilistic association index p-ASI which corresponds to the ratio of A_m cells that are spatially associated to A_l cells:

$$\text{pASI}(A_l, A_m) = \frac{\eta^l_a(A_m)}{|A_m|} \in [0;1], \tag{20}$$

with $\eta^l_a(A_m) = \sum_{i=1}^{n}\left\{\eta^l_a(A_m)\right\}_{\omega_i}$ the total number of associated A_m cells.

We also derived the probability-weighted coupling distance

$$\delta_a(A_l, A_m) = \frac{1}{|A_m|}\sum_{j=1}^{N}\sum_{i=1}^{n}\chi_i(u_j)p^m_j\left(\frac{\{\eta^l_a(A_m)\}_{\omega_i}}{\eta^l_{\omega_i}}\right)|u_j - \partial A_l|, \tag{21}$$

where the ratio $\frac{\{\eta^l_a(A_m)\}_{\omega_i}}{\eta^l_{\omega_i}}$ is the proportion of A_m cells within ω_i that are spatially associated to A_l cells. It is the probability that cell u_j is spatially associated and not randomly distributed, and $|u_j - \partial A_l|$ is the Euclidean distance of u_j to A_l cells' contours. Finally, the association p-value is given by

$$p - value = 1 - \psi^n\left(\sup_{1 \le i \le n}\frac{K^l_{\omega_i}(A_m) - \mu^l_i}{\sigma^l_i}\right), \tag{22}$$

where $\psi(.)$ is the cumulative density function of a normalized Gaussian random variable.

Validation levelset analysis with synthetic simulations

To validate the accuracy of DeepCellMap in characterizing the spatial association between two cell types, we generated synthetic simulations with various spatial distributions. For the sake of simplicity, we considered three cell types: A, B, and C, and cells were reduced to points. Positions were uniformly distributed in a square domain Ω. Among the B and C positions, a part $n^{Random}_B = n_B \times (1 - C_B)$ (resp. $n^{Random}_C = n_C \times (1 - C_C)$) of cells are randomly distributed in Ω, while the remaining part $n^{Gaussian}_B = n_B \times C_B$ (resp. $n^{Gaussian}_C = n_C \times C_C$) are positioned around A cells with a Gaussian association distance $d \sim \mathcal{N}(\mu_B, \sigma_B)$ (resp. $d \sim \mathcal{N}(\mu_C, \sigma_C)$). To efficiently sample the coupled B and C cells, the distance transform from A cells is computed across the entire image, a random association distance d is drawn from $\mathcal{N}(\mu, \sigma)$, and a coordinate in the distance map at distance d from A cells is randomly selected (Supplementary Fig. 11). To assess the capability of DeepCellMap to handle cell misclassification, we used three different confusion scenarios: scenario P1 (no confusion), there are no classification errors, which allows the validation of the analysis of spatial association in a setting where there is no ambiguity in cell types; scenario P2 (intermediate) were 30% of B cells are classified as C cells and vice-versa; scenario P3 (high) where classification errors reach 45% of classification errors between types B and C. Parameters used for the simulations are summarized in the Supplementary Material.

The accuracy of DeepCellMap was measured by comparing the estimated association with ($\hat{C}^c_i$) and without ($\hat{C}_i$) correcting for misclassification, with ground truth C_i between cells of type $i \in [B, C]$ and cells A. The estimated association distances $\hat{\delta}_i$ and $\hat{\delta}^c_i$ were compared to μ_i.

Characterizing cell clustering and overlap with a generalized DBSCAN algorithm

To measure the rate of clustering of the different cell types, and the overlap between the domains occupied by the different cell populations, we developed an improved and automatically-calibrated *density-based spatial clustering of applications with noise* (DBSCAN) algorithm[26]. This widely used clustering algorithm contains two user-defined parameters: the radius ϵ to define the neighborhood of each cell and the number Min_{Sample} which defines the minimum number of neighboring cells (points) in this neighborhood in order to consider the cell as belonging to a cluster. We used the DBSCAN algorithm with cell coordinates (x, y) for each microglial state (Fig. 4A) and automatically compute ϵ to maximize the number of clusters (Fig. 4B):

$$\epsilon^* = \text{Argmax}_\epsilon(\#\text{cluster}(\varepsilon)). \tag{23}$$

We observed empirically that the number of clusters #cluster(ε) has a single maximum (Fig. 4C), ensuring that our selection procedure converges. We fixed the minimum number *MinSample* = 4 of points inside a cell neighborhood (disk) to classify it as *clustered*.

To ensure the stability of computed clusters, we defined the following criteria: after removing 10% of the cells located on the cluster's convex hull[65], we computed the ratio between the area of the remaining cluster and the initial one and kept the cluster if this ratio exceeds a threshold $t = 60\%$ (average over 100 realizations). After having sorted each cell as part of a cluster or not (isolated) we computed several metrics to quantify cell clustering and the relative overlap between the clusters from different types: the fraction of clustered cells (*versus* isolated) (Fig. 4F), the mixing proportion $\phi_{A/B}$ between clusters of cells from type A and B. This mixing proportion is computed as the fraction of A clustered cells belonging to the convex hull of B clusters (Fig. 4H):

$$\phi_{A/B} = \frac{1}{n_A^c} \sum_{i=1}^{n_A^c} \chi_B(\text{cell}_i), \tag{24}$$

where n_A^c is the number of *A* clustered cells and $\chi_B(\text{cell}_i) = 1$ if $\text{cell}_i \in B$ clusters convex hull and 0 otherwise.

Cell neighbors analysis

DeepCellMap allows for a traditional quantification of the relationships between each cell's neighbors. We used various parameters to estimate the relative position of different cell populations in a region of interest:

1. For cells of type A, we define the first neighbor distance as

$$d_1^A = \frac{1}{n_A} \sum_{i=1}^{n_A} d_1^i, \tag{25}$$

where n_A is the total number of cells A and d_1^i are the distances to the first neighbors of cell i (regardless of neighbor type).

2. For A cells, we define the distance to the first B neighbor as

$$d_{B,1}^A = \frac{1}{n_A} \sum_{i=1}^{n_A} d_{B,1}^i, \tag{26}$$

where $d_{B,1}^i$ is the distance to the first *B* neighbor of cell *i*. This distance $d_{B,1}^A$ can be interpreted as the average distance from one population to another. For these last two metrics, DeepCellMap includes the possibility to calculate d_k^A and $d_{B,k}^A$, $\forall k$.

Reporting summary

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

Data availability

Most of the cases are part of the Oxford Brain Bank and Zagreb Brain Bank digital archives and will be made accessible for research under a cost recovery model, contact david.menassa@ndcn.ox.ac.uk to arrange access. HDBR cases can be requested directly by contacting katie.long@kcl.ac.uk. Requests for access will be dealt with immediately and fulfilled within 2 weeks. Source data are provided in this paper.

Code availability

Codes are available in Zenodo of the Holcman's lab DOI with URL https://doi.org/10.5281/zenodo.14003597 and GitHub (https://github.com/holcman-lab/DeepCellMap) and also Bionewmetrics for a description at www.Bionewmetrics.org.

References

1. Fischer, A. H., Jacobson, K. A., Rose, J. & Zeller, R. Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harb. Protoc.* **2008**, pdb–prot4986 (2008).
2. Skoog, L. & Tani, E. Immunocytochemistry: an indispensable technique in routine cytology. *Cytopathology* **22**, 215–229 (2011).
3. Black, S. et al. Codex multiplexed tissue imaging with dna-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
4. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
5. Menassa, D. A. et al. The spatiotemporal dynamics of microglia across the human lifespan. *Dev. Cell* **57**, 2127–2139 (2022).
6. Masuda, T. et al. Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* **566**, 388–392 (2019).
7. Menassa, D. A., Kopić, J., Junaković, A., Kostović, I. & Krsnik, Ž. Microglial characterization in transient human neurodevelopmental structures. *Dev. Neurosci.* **45**, 1–7 (2023).
8. Menassa, D. A. & Gomez-Nicola, D. Microglial dynamics during human brain development. *Front. Immunol.* **9**, 1014 (2018).
9. Barry-Carroll, L. et al. Microglial colonisation of the developing brain is facilitated by clonal expansion of highly proliferative progenitors and follows an allometric scaling. *Cell Rep.* **42**, 112425 (2023).
10. Ratz, M. et al. Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat. Neurosci.* **25**, 285–294 (2022).
11. Davalos, D. et al. Atp mediates rapid microglial response to local brain injury in vivo. *Nat. Neurosci.* **8**, 752–758 (2005).
12. Han, W. et al. Cranial irradiation induces transient microglia accumulation, followed by long-lasting inflammation and loss of microglia. *Oncotarget* **7**, 82305 (2016).
13. Kettenmann, H., Hanisch, U.-K., Noda, M. & Verkhratsky, A. Physiology of microglia. *Physiol. Rev.* **91**, 461–553 (2011).
14. Massimo, M. et al. Haemorrhage of human foetal cortex associated with sars-cov-2 infection. *Brain* **146**, 1175–1185 (2023).
15. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, 265–273 (Springer, 2018).
16. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
17. Stetzik, L. et al. A novel automated morphological analysis of iba1+ microglia using a deep learning assisted model. *Front. Cell. Neurosci.* **16**, 944875 (2022).
18. Colombo, G. et al. A tool for mapping microglial morphology, morphomics, reveals brain-region and sex-dependent phenotypes. *Nat. Neurosci.* **25**, 1379–1393 (2022).
19. Davis, B. M., Salinas-Navarro, M., Cordeiro, M. F., Moons, L. & De Groef, L. Characterizing microglia activation: a spatial statistics approach to maximize information extraction. *Sci. Rep.* **7**, 1576 (2017).
20. Leyh, J. et al. Classification of microglial morphological phenotypes using machine learning. *Front. Cell. Neurosci.* **15**, 701673 (2021).
21. Li, X. et al. Infiltration of cd8+ t cells into tumor cell clusters in triple-negative breast cancer. *Proc. Natl. Acad. Sci. USA* **116**, 3678–3687 (2019).
22. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182**, 1341–1359 (2020).
23. Lagache, T., Lang, G., Sauvonnet, N. & Olivo-Marin, J.-C. Analysis of the spatial organization of molecules with robust statistics. *PLOS ONE* **8**, 1–7 (2013).

24. Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns* (John Wiley & Sons, 2008).

25. Lagache, T., Sauvonnet, N., Danglot, L. & Olivo-Marin, J. Statistical analysis of molecule colocalization in bioimaging: Statistical analysis of molecule colocalization in bioimaging. *Cytometry Part A* **87**, 568–579 (2015).

26. Ester, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 226–231 (1996).

27. Kerwin, J. et al. The hudsen atlas: a three-dimensional (3d) spatial framework for studying gene expression in the developing human brain. *J. Anat.* **217**, 289–299 (2010).

28. Kettenmann, H., Hanisch, U. K., Noda, M. & Verkhratsky, A. Physiology of microglia. *Physiol. Rev.* **91**, 461–553 (2011).

29. Vidal-Itriago, A. et al. Microglia morphophysiological diversity and its implications for the cns. *Front. Immunol.* **13**, 997786 (2022).

30. Streit, W. J., Braak, H., Xue, Q. S. & Bechmann, I. Dystrophic (senescent) rather than activated microglial cells are associated with tau pathology and likely precede neurodegeneration in alzheimer's disease. *Acta Neuropathol.* **118**, 475–85 (2009).

31. Bisht, K. et al. Dark microglia: A new phenotype predominantly associated with pathological states. *Glia* **64**, 826–39 (2016).

32. Hornik, T. C., Neniskyte, U. & Brown, G. C. Inflammation induces multinucleation of microglia via pkc inhibition of cytokinesis, generating highly phagocytic multinucleated giant cells. *J. Neurochem.* **128**, 650–61 (2014).

33. Martín-Estebané, M. et al. Onset of microglial entry into developing quail retina coincides with increased expression of active caspase-3 and is mediated by extracellular atp and udp. *PLOS ONE* **12**, e0182450 (2017).

34. Summers, H. D., Wills, J. W. & Rees, P. Spatial statistics is a comprehensive tool for quantifying cell neighbor relationships and biological processes via tissue image analysis. *Cell Rep. Methods* **2**, https://doi.org/10.1016/j.crmeth.2022.100348 (2022).

35. Tao, Y. et al. Cntools: A computational toolbox for cellular neighborhood analysis from multiplexed images. *PLOS Comput. Biol.* **20**, e1012344 (2024).

36. Ruitenberg, M. J. & Nguyen, Q. H. Cellular neighborhood analysis in spatial omics reveals new tissue domains and cell subtypes. *Nat. Genet.* **56**, 362–364 (2024).

37. Canete, N. P. et al. spicyr: spatial analysis of in situ cytometry data in r. *Bioinformatics* **38**, 3099–3105 (2022).

38. Mukherjee, S., Gomez, C., Danglot, L., Lagache, T. & Olivo-Marin, J.-C. Generalizing the statistical analysis of objects' spatial coupling in bioimaging. *IEEE Signal Process. Lett.* **PP**, 1–1 (2020).

39. Lagache, T. et al. Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nat. Commun.* **9**, 698 (2018).

40. Sharma, K., Bisht, K. & Eyo, U. B. A comparative biology of microglia across species. *Front. Cell Dev. Biol.* **9**, 652748 (2021).

41. Barry-Carroll, L. & Gomez-Nicola, D. The molecular determinants of microglial developmental dynamics. *Nat. Rev. Neurosci.* **25**, 414–427 (2024).

42. Askew, K. et al. Coupled proliferation and apoptosis maintain the rapid turnover of microglia in the adult brain. *Cell Rep.* **18**, 391–405 (2017).

43. Cunningham, C. L., Martínez-Cerdeño, V. & Noctor, S. C. Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *J. Neurosci.* **33**, 4216–4233 (2013).

44. Xu, J., Wang, T., Wu, Y., Jin, W. & Wen, Z. Microglia colonization of developing zebrafish midbrain is promoted by apoptotic neuron and lysophosphatidylcholine. *Dev. Cell* **38**, 214–222 (2016).

45. Kang, R. et al. The dual role of microglia in blood-brain barrier dysfunction after stroke. *Curr. Neuropharmacol.* **18**, 1237–1249 (2020).

46. Jellinger, K. A. & Korczyn, A. D. Are dementia with lewy bodies and parkinson's disease dementia the same disease? *BMC Med.* **16**, 1–16 (2018).

47. Sieben, C., Banterle, N., Douglass, K. M., Gönczy, P. & Manley, S. Multicolor single-particle reconstruction of protein complexes. *Nat. Methods* **15**, 777–780 (2018).

48. Lewis, S. M. et al. Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods* **18**, 997–1012 (2021).

49. Gorris, M. A. et al. Eight-color multiplex immunohistochemistry for simultaneous detection of multiple immune checkpoint molecules within the tumor microenvironment. *J. Immunol.* **200**, 347–354 (2018).

50. Ali, H. R. et al. Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nat. Cancer* **1**, 163–175 (2020).

51. Del Bigio, M. R. Cell proliferation in human ganglionic eminence and suppression after prematurity-associated haemorrhage. *Brain* **134**, 1344–1361 (2011).

52. Onorati, M. et al. Molecular and functional definition of the developing human striatum. *Nat. Neurosci.* **17**, 1804–15 (2014).

53. Torres-Platas, S. G. et al. Morphometric characterization of microglial phenotypes in human cerebral cortex. *J. Neuroinflammation* **11**, 12 (2014).

54. Braun, E. et al. Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).

55. Kracht, L. et al. Human fetal microglia acquire homeostatic immune-sensing properties early in development. *Science* **369**, 530–537 (2020).

56. 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).

57. Butovsky, O. & Weiner, H. L. Microglial signatures and their role in health and disease. *Nat. Rev. Neurosci.* **19**, 622–635 (2018).

58. Sierra, A., Abiega, O., Shahraz, A. & Neumann, H. Janus-faced microglia: beneficial and detrimental consequences of microglial phagocytosis. *Front. Cell. Neurosci.* **7**, 6 (2013).

59. Silva-Gomes, R. et al. Differential expression and regulation of ms4a family members in myeloid cells in physiological and pathological conditions. *J. Leukoc. Biol.* **111**, 817–836 (2022).

60. Boghozian, R., Sharma, S., Narayana, K., Cheema, M. & Brown, C. E. Sex and interferon gamma signaling regulate microglia migration in the adult mouse cortex in vivo. *Proc. Natl. Acad. Sci. USA* **120**, e2302892120 (2023).

61. Ruifrok, A. C. et al. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**, 291–299 (2001).

62. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).

63. Mukherjee, S. et al. Spatial analysis for histopathology: A statistical approach. In *2022 IEEE 19th International Symposium on Biomedical Imaging* 1–5 (2022).

64. Donoho, D. L. & Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455 (1994).

65. Gillies, S. et al. Shapely: manipulation and analysis of geometric objects https://github.com/Toblerity/Shapely (2007).

Acknowledgements

Author contributions

D.H. co-supervised T.P. with D.A.M. and T.L. D.H., T.L., and D.A.M. designed the overall project. D.A.M., L.P., and Z.K. collected fetal human tissues, labeled them with immunohistochemistry, and digitized them for computational analyses. K.L. and M.M. provided the SARS-CoV-2 dataset. X.L. and E.M. performed the sc-RNA-seq analysis. Y.R., K.B., and A.L.R. analyzed datasets for further profiling including spatial transcriptomic datasets. T.P. wrote the code. T.P., T.L., and D.H. developed the method and analyzed the data. All authors wrote and revised the manuscript extensively. Authorship was granted as per the guidelines provided by the journal.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-56560-z.

Correspondence and requests for materials should be addressed to Thibault Lagache, David A. Menassa or David Holcman.

Peer review information *Nature Communications* thanks Roman Sankowski, Jia Song, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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

[1]Group of Data Modeling and Computational Biology, IBENS, École Normale Supérieure, Paris, France. [2]Croatian Institute for Brain Research, University of Zagreb, Zagreb, Croatia. [3]Centre for Developmental Neurobiology, MRC Centre for Neurodevelopmental Disorders, King's College London, London, UK. [4]Department of Women's and Children's Health, Karolinska Institutet, Solna, Sweden. [5]Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Solna, Sweden. [6]Department of Neuropathology and The Queen's College, University of Oxford, Oxford, United Kingdom. [7]BioImage Analysis Unit, CNRS UMR3691, Institut Pasteur, Université Paris Cité, Paris, France. [8]DAMPT, University of Cambridge, DAMPT and Churchill College, Cambridge, United Kingdom. [9]These authors jointly supervised this work: Thibault Lagache, David A Menassa, David Holcman. ✉e-mail: thibault.lagache@pasteur.fr; david.menassa@queens.ox.ac.uk; david.holcman@ens.psl.eu