Supplementary Information

Title:
Senescent and disease-associated microglia are modifiable features of aged brain white matter

AUTHORS and AFFILIATIONS:
Chase M. Carver[1,2]
Paul T. Gomez[1,2]
Sonia L. Rodriguez[1,2]
Jennifer M. Kachergus[3]
Yi Liu[4]
Ji Shi[3]
Tommy Tran[5]
Liguo Wang[6]
Simon Melov[5]
E. Aubrey Thompson[3]
Marissa J. Schafer[1,2,7]*

[1]Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, USA
[2]Robert and Arlene Kogod Center on Aging, Mayo Clinic, Rochester, MN, USA
[3]Department of Cancer Biology, Mayo Clinic, Jacksonville, FL, USA
[4]Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, FL, USA
[5]Buck Institute for Research on Aging, Novato, CA, USA
[6]Division of Computational Biology, Mayo Clinic College of Medicine, Rochester, MN, USA
[7]Department of Neurology, Mayo Clinic, Rochester, MN, USA

*Corresponding Author

Supplementary Text.

CosMx raw data pipeline. CosMx raw transcript localizations, cell segmentation polygons, and FOV image data were initially processed through the Giotto Suite R package (3.3.0). FOVs to be analyzed from a single brain cross-section were compiled into a list and transcripts and polygons from those FOVs were read into data frames from the CosMx-generated csv files.

Cell segmentation and edge-filtering. A data frame of cell segmentation polygons was created and read as "Well-known text" (WKT) representation of geometry format. *A priori*, a 50-pixel boundary was determined for each FOV at the top and right edges. Each polygon was processed through a filtering algorithm and excluded from downstream analysis if it intersected with either the top or right edge boundary. Polygons that passed the boundary-exclusion algorithm were assembled into a new cell segment data frame. Only transcripts that were contained within the new segments were analyzed.

Giotto object of joined FOVs. The cell overlay images for each FOV were composited into an image containing all FOVs listed for analysis. Gene feature transcripts and negative transcript coordinates were tabulated into cell segments from the FOVs. A subset object of transcripts, segments, and cell overlay images were created for each FOV: createGiottoPolygonsFromDfr(), createGiottoObjectSubcellular(), addSpatialCentroidLocations(), addGiottoImage(), convertGiottoLargeImageToMG(). FOV subsets were joined into a single Giotto object file for downstream analyses: joinGiottoObjects(). The feature points overlapping onto cell segment polygons were returned to the relevant giottoPolygon object's overlap slot: calculateOverlapRaster(). The overlap information was converted into a cell by feature expression matrix, which was then stored in the Giotto object's expression slots: overlapToMatrix().

Expression slot processing. The cells within the FOV-joined Giotto object were filtered by excluding cells with less than 20 features detected: filterGiotto(). The raw expression values were used at an expression threshold of 1. The Giotto object was then normalized by "standard" method as provided by Giotto Suite documentation, normalizeGiotto(), addStatistics(): data normalization for total library size and scaling by a custom scale-factor (5000), log2 transformation of data, and Z-scoring of data by genes and or cells. Separately for dimensionality reduction to perform single-cell clustering, the Giotto object was also normalized by Lause/Kobak method by calculating the expected values based on Pearson correlations, and calculating z-scores based on observed and expected values. The highly variable gene features were then computed from the Pearson expression values, calculateHVF(), and used to create a principal component analysis, runPCA(). A UMAP was generated from the PCA with 1:10 dimensions, 8 n_neighbors, 3 n_components, 0.1 min_dist, and 4 threads: runUMAP(). A nearest-neighbor network was then created with 1:20 dimensions, and k of 10: createNearestNetwork(). Leiden clustering followed with a resolution of 0.10-0.11 across 1000 n_iterations: doLeidenCluster(). The microglia cells were isolated as a dataset by expression of based on previously validated markers *Csf1r*, *Ctss*, *Cx3cr1*, *Hexb*, *Selplg*, *Itgam*, *P2ry12, Tmem119*, and *Trem2* (Zeisel et al., 2018). Macrophages were delineated from the myeloid-lineage cells of the brain based on expression of key markers *Crip1, Cd74, Fxyd5, H2-Aa, H2-Ab1, H3f3b* (**Supplementary Figure 1**, below). Markers for each microglia cluster were determined by Gini coefficient or Scran methods using findMarkers_one_vs_all(), using the leiden cluster information as the cluster_column.

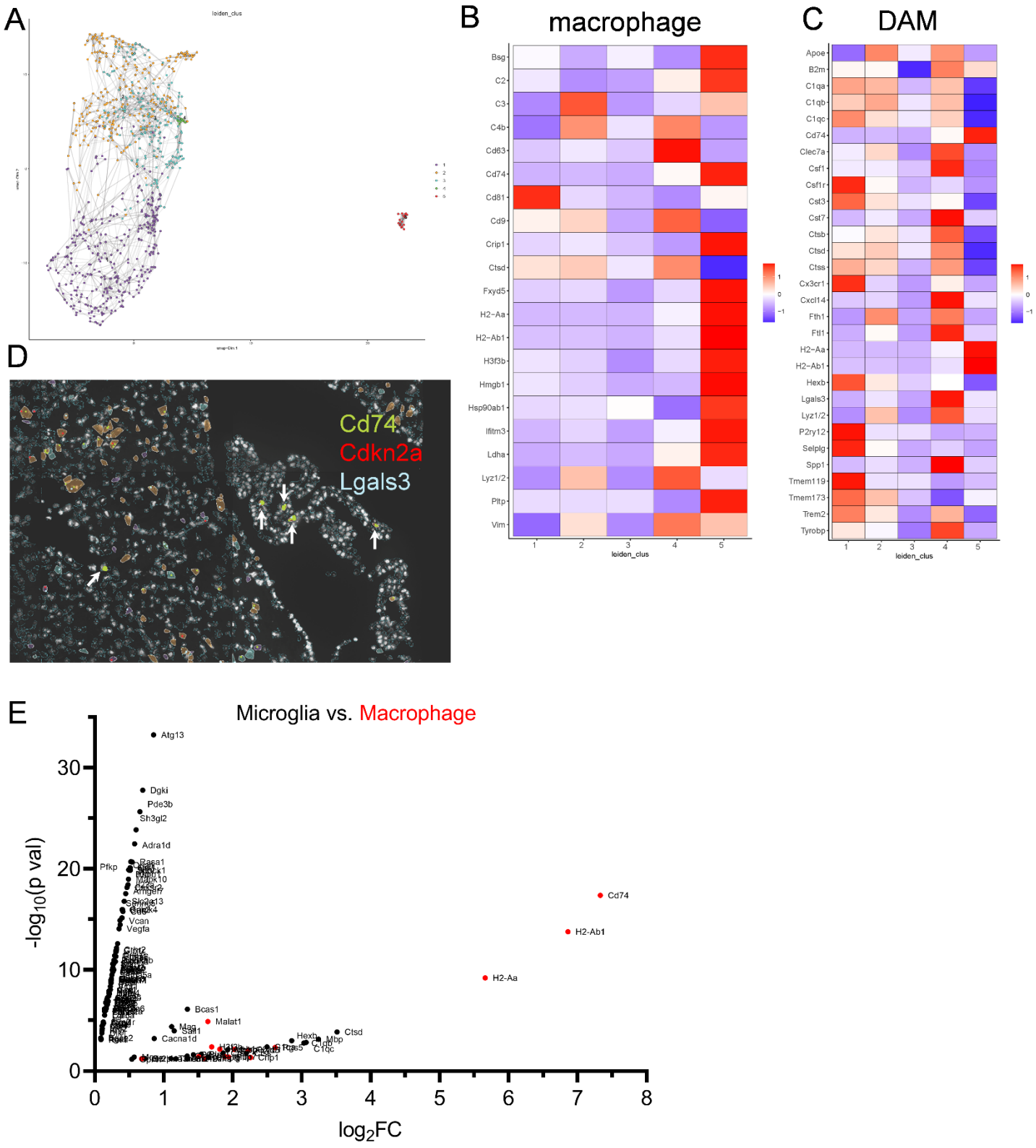
Zeisel A*, et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e1022 (2018)

Seurat pooled cluster processing. For each brain section Giotto object, a microglia Seurat cluster was created using the giottoToSeurat() function. Each Seurat object received an updated metadata annotation to include the brain of origin #231-234. A pooled brain microglia Seurat object was created by merging the datasets. The pooled object's data was scaled using all gene features: ScaleData(), followed by finding the variable features to run PCA: FindVariableFeatures(), followed by finding nearest neighbors FindNeighbors() and generating clusters with Leiden algorithm as 0.5 resolution, FindClusters(). A UMAP was created using the first 10 dimensions of the PCA analysis: RunUMAP().

Specific gene-enriched cell comparisons. Microglia cell populations enriched for *Lgals3* gene or cyclin-dependent kinase inhibitor genes were subset from the pooled Seurat object based on a count of 2 or more transcripts per

cell. This enriched subset was then compared to a subset of all other microglia using the Wilcoxon Rank Sum test to identify differentially expressed genes with the FindMarkers() function.

Information for access to raw data. Analyses data can be found in the Extended Data Tables included with this article. Spatial transcriptomics raw datasets will be published through the NIH's Cellular Senescence Network Consortium Organization and Data Coordinating Center (CODCC) upon publication of this manuscript.



Supplementary Figure 1. CosMx analysis of myeloid-lineage cells including microglia and macrophages. (A) UMAP, Leiden clustering, and nearest-neighbor network from brain ID #231, which includes cluster 5, a macrophage-specific population far-removed from all other microglia in clusters 1-4, and therefore, was removed from analyses described in Figure 4, wherein, distinct cluster profiles are described. (B) Z-score heatmap of macrophage-specific genes (normalized Gini coefficient). Cluster 5 shows relative high expression of genes *Crip1, Cd74, Fxyd5, H2-Aa, H2-Ab1, H3f3b*, and low expression of *Cd9* and *Ctsd*. (C) Z-score heatmap of

disease-associated microglia (DAM) genes (normalized Gini coefficient). Cluster 5 exhibits lowest relative expression of *C1qa, C1qb, C1qc, Cst3, Ctsb, Ctsd, Ctss, Hexb, and Trem2*. (D) CosMx transcript localizations overlayed onto microglia and macrophage (arrows) segmented polygons. Transcripts for *Cd74* (green)*, Cdkn2a* (red), and *Lgals3* (blue) are shown, in which macrophages in choroid plexus and brain are identified by dense *Cd74* transcripts. (E) Volcano plot of $\log_2$ fold-changes of genes depicting relative high expression in microglia (black dots) and relative high expression in macrophages (red).