nature neuroscience

Article

https://doi.org/10.1038/s41593-024-01573-y

Proteomic and transcriptomic profiling of brainstem, cerebellum and olfactory tissues in early- and late-phase COVID-19

In the format provided by the authors and unedited

Supplementary Material

Additional Material and Methods

Autopsy Sample collection

All COVID-19 patients died from SARS-CoV-2 pneumonia, respiratory failure, or terminal multiorgan dysfunction syndrome as result of bacterial superinfection and sepsis. We selected patients with a short post-mortem interval (PMI) and available cryo-tissue. Patients with concomitant autoimmune diseases, tumors, or infarcts/bleedings in the selected regions were excluded. We summarized the comorbidities and clinical data of donors and the respective analysis done from which sample in **Supplementary table 1**. Cerebellum was taken of the cerebellar cortex; the brainstem sample was taken at the level of the inferior olive and the hypoglossal nucleus. Lung samples are from a previous publication[1]. For all tissues/regions samples were histologically checked regarding structural pathology and avoiding infarcts, abscesses and hemorrhages. The clinical records were assessed thoroughly for pre-existing medical conditions, medications and progression of the disease as well as COVID-19-related symptoms before death. All autopsies were performed from spring 2020 to summer 2021.

CSF sample collection

For the cerebrospinal fluid (CSF) dataset we performed a retrospective study on COVID-19 patients with lumbar puncture to rule out CNS pathologies such as autoimmune encephalitis or meningitis involving a total *n* = 38. Mean age was 68.6 years with 28 male and 9 female patients. The main indications for performing lumbar puncture were encephalopathy or prolonged awakening after weaning. SARS-CoV-2 patients were divided into Covid-19lowPCT and Covid-19highPCT and compared to *n* = 10 HSV meningitis patients and *n* = 11 non-inflammatory controls (**Supplementary table 1, tab CSF cohort**).

snRNA Sequencing

The single-nucleus postmortem tissue dataset was processed using cellranger 3.0.1. All transcripts were aligned to a customized hg19 reference transcriptome (10x Genomics, version 3.1.0) that included the SARS-CoV-2 genome (Refseq-ID: NC_045512) as an additional chromosome. Ambient RNA removal was performed for each sample with SoupX 1.4.5[2]. Further pre-processing was performed with Seurat 3.1.4[3,4] where cells with more than or equal to 10% mitochondrial reads or less than 200 genes expressed were filtered out. An upper cutoff for the number of genes was selected manually for each sample based on the

outliers in a UMI counts vs. gene counts plot (3,000-6,000 genes). After log-normalization, samples were integrated with canonical correlation analysis (CCA) in Seurat. Using the integrated dataset, PCA was run, followed by UMAP calculation and finally clustering. Cell type assignment was performed as previously described and by using marker genes[1,5-9]. Differentially expressed genes for each cluster are provided in **Supplementary table 5**. Cell numbers and median genes are also given in **Supplementary table 5**. A cluster of cells were left undefined that did not express known marker genes of brain cell types. Microglia/macrophage cluster 1 shows less expression of *PTPRC (CD45), CD47, ITGAM (CD11b), CD14* and *CD68* compared to macrophage/microglia cluster2 and a higher expression of *P2YR12* and *TMEM119*, so we assume that cluster 1 contains more 'microglia' and cluster2 more 'macrophages'[10] (**Extended Data Figures 8 a, d**).

In the medulla dataset endothelial cell cluster 1 shares features of venous endothelial cells, cluster 2 of arterial endothelial cells and endothelial cluster 3 of capillary endothelial cells[11] and were named accordingly. The neuronal clusters show no exclusive neurotransmitter profiles, however neuron cluster 1 and 2 show enriched expression of genes related to dopaminergic neurons and neuron cluster 3 of glutamatergic neurons[12-19] (**Extended Data Figures 8 c, e**).

Statistical analysis for gene expression of snRNA Sequencing

When two groups were compared, the student's t-test was applied. For three or more group comparisons, one or two-way analysis of variance (ANOVA) was performed, followed by Tukey's or Bonferroni's post hoc test, as appropriate. All t-test t- and *p*-values and ANOVA F and *p*-values are reported in the legends; significant results of the contrast tests are further indicated by asterisks in the graphs. All statistical tests were two-tailed and $p < 0.05$ was considered statistically significant. For the complete set of raw data, see the data source file. Differential gene expression testing was calculated using FindMarkers with the MAST-based differential expression test. *P*-values were adjusted using the Benjamini-Hochberg method. Significance was calculated for IFN scores with a Mann-Whitney U test where the *p*-values were corrected with the Benjamini-Hochberg method. This is followed by a Wilcoxon effect size calculation for significant differences only. Significant differences in cellular frequencies were calculated with scCODA, a Bayesian model that determines compositional changes in single-cell RNA-seq datasets (significant if FDR adjusted *p*-value < 0.1)[20] and are reported in **Supplementary table 7.**

Multiplex histology

The donors analyzed are specified in **Supplementary table 1**. We analyzed *n* = 2 late and *n* = 2 control cases. 5 µm cryo-tissue sections were cut using a MH560 cryotome (ThermoFisher, Waltham, Massachusetts, USA) and mounted on coverslips (24 x 60 mm; Menzel-Gläser, Braunschweig, Germany) that had been coated with 3-aminopropyltriethoxysilane (APES). The sections were fixed for 8 minutes at room temperature (RT) with 2% paraformaldehyde (methanol- and RNAse-free; Electron Microscopy Sciences, Hatfield, Philadelphia, USA), washed, and permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature. Afterwards, sections were blocked using 10% goat serum and 5% BSA in PBS for 20 minutes. For multiplex histology, a fluid chamber for staining and washing solution was created using “press-to-seal” silicone sheets (Life technologies, Carlsbad, California, USA; 1.0 mm thickness) with a 10 mm circular cut-out, which was glued firmly to the coverslip surrounding the sample.

Multiplex microscopy image acquisition was performed as previously published[21,22]. In short, we used a modified Toponome Image Cycler® MM3 (TIC) originally produced by MelTec GmbH & Co.KG Magdeburg, Germany[23] to generate multiplex histology data. The ImageCycler is a robotic microscopic system with 3 main components: (1) an inverted widefield (epi)fluorescence microscope (Leica DM IRE2) equipped with a CMOS camera and a motor-controlled XY-stage, (2) CAVRO XL3000 Pipette/Diluter (Tecan GmbH, Crailsheim, Germany), and (3) M for controlling microscope and pipetting system and for automated image acquisition. The multiplex microscopy experiment is a sequence of cycles, each containing four steps: (i) pipetting of the fluorescence-coupled antibody staining solution onto the sample, followed by incubation and subsequent washing; (ii) auto-focusing based on cross correlation of phase contrast images, followed by acquisition of the fluorescence images 3D stack (+/- 5 z-steps); (iii) photo-bleaching of the fluorophore; and (iv) a second auto-focusing step followed by acquisition of a post-bleaching fluorescence image 3D stack (+/- 5 z-steps). Afterwards, the cycle is repeated with the next fluorescently labeled antibody. After the sample was labeled by all antibodies of interest as described above, the experiment is completed. The antibodies and respective dilutions used for multiplexed immunofluorescence histology of the medulla sample are listed in **Table 1**. The antibodies were stained in the indicated order.

All images were aligned by cross-correlation based on the reference phase contrast image taken at the beginning of the measurement. Afterwards, each fluorescence MELC image was processed by background subtraction and illumination correction, based on the signal of the bleaching images[23]. In order to account for slice thickness, an “Extended Depth of Field” algorithm was applied on the 3D fluorescence stack in each cycle[24]. Images were then

normalized in Fiji[25], where a rolling ball algorithm was used for background estimation, edges were removed (accounting for the maximum allowed shift during the autofocus procedure) and fluorescence intensities were stretched to the full intensity range (16 bit => 2^{16}). The 2D fluorescence images generated in this way were subsequently analyzed using Fiji/ImageJ (version 1.52i).

Table 1. Antibodies used for MELC.

antibodies	source	identifier	dilution	incubation time (s)
DAPI	Roche	Cat# 10236276001, N/A	´1:5000	1800
Laminin	Abcam	ab11575	´1:200	1800
goat-anti-rabbit-A555	Invitrogen	A21428	´1:200	1800
CD3-PE	Miltenyi Biotec	Cat# 130-113-139	´1:50	2700
CD4-PE	Miltenyi Biotec	Cat# 130-113-214, RRID:AB_2726025	´1:50	2700
CD8-PE	Miltenyi Biotec	Cat# 130-113-720, RRID:AB_2726261	´1:50	2700
PD1-PE	Miltenyi Biotec	Cat# 130-120-388, RRID:AB_2752074	´1:50	2700
CD163	Biolegend	326505	´1:50	2700
CD49a-PE	Miltenyi Biotec	Cat# 328304, RRID:AB_1236407	´1:50	2700
Granzyme B-PE	Miltenyi Biotec	Cat# 130-116-654, RRID:AB_2727639	´1:50	2700
CD45-PE	Miltenyi Biotec	Cat# 130-113-118, RRID:AB_2725946	´1:50	2700
CD31-PE	R and D Systems	Cat# FAB3567P, RRID:AB_2279388	´1:50	1800
HLA-DR,DP,DQ-PE	Miltenyi Biotec	Cat# 130-120-715, RRID:AB_2752176	´1:50	1800
GFAP	Dako	Z 0334	´1:1000	1800
SMA-FITC	Abcam	Cat# ab8211, RRID:AB_306359	´1:50	1800

Immunohistochemical procedures

The samples analyzed with the respective staining are specified in **Supplementary table 1**. 4-µm-thick FFPE or 8-µm-thick frozen tissue sections were used for additional immunohistochemical stainings which were performed manually or on a Benchmark XT autostainer (Ventana Medical Systems, Tuscon, AZ, USA) with standard antigen retrieval methods (CC1 buffer, pH8.0, Ventana Medical Systems, Tuscon, AZ, USA). The following primary and secondary antibodies were used: Anti-SARS-CoV-2 antibody (Synaptic Systems) 1:2000 with Ventana Benchmark XT autostainer (Ventana, Tuscon, Arizona, USA) or rabbit anti–pStat1 (Y701) (clone D4A7,1:1000, heat mediated antigen retrieval with TrisEDTA, pH 8.0, cell signaling #7649). Manual pStat1 staining was developed with the Envision+ System for mouse and rabbit (Dako) and stainings were digitized using a whole slide scanner (Olympus VS120 or Leica Aperio 450 DX). Image visualization was performed using Omero Server software 5.6, Omero figure 4.0.2 (https://github.com/ome/omero-figure).

All staining were developed using the Ultra View Universal 3,3´-Diaminobenzidine (DAB) Detection Kit (Ventana, Roche) which contains both secondary antibodies (anti-rabbit and anti-mouse), DAB stain and counter staining reagent.

IHC stained sections were evaluated by at least two independent experienced consultant Neuropathologists.

Interferome database

To reliably identify Interferon regulated genes (IRGs) and IRG signatures, we used Interferome v2.01[26] (http://www.interferome.org/interferome/home.jspx). The search conditions and results for each anatomical region are given in **Supplementary table 2**. To generate the IFN score, all differentially upregulated genes in COVID-19 acute compared to the control (logFC > 0.25) were obtained and parsed through the Interferome database to identify IRGs. This IFN genelist was then used to calculate a score using the AddModuleScore function in Seurat[27]. This generates an average expression for the IFN program that is subtracted by the aggregate expression of control genes.

Synapse database

Synaptic gene identification was done using the genomic analysis tool SynGO (https://syngoportal.org/index.html)[28], an open-access expert-curated dataset of synaptic genes. The differentially expressed gene sets of our cohort were uploaded to SynGo[28]. Preset settings were retained with a default brain-expressed gene-list, defined as 'expressed in any GTEx v7 brain' as a reference set (**bulk download SynGO release 1.1.**). Results were further analyzed using the annotated GO terms **(Supplementary table 4)**.

Proteomics

Data pre-processing

Each tissue type was processed separately. DIA_NN normalized peptide intensities were used as input. Samples with low quality were removed. Peptides with excessive missing values (> 40% per group) were excluded from analysis. To the rest of peptides group-based imputation of missing values using PCA method[29] was applied. Then they were normalized using vsn method[30]. To obtain a quantitative protein data matrix, the log2-intensities of peptides were filtered, only peptides belonging to one protein group were kept, and then summarized by “maxLFQ” method[31], implemented in R package iq[32] into protein log intensity. Change of the data matrix after certain pre-processing steps is presented in **Table 2**.

Table 2. Change of the data matrix size after pre-processing steps.

Bulbus		N features	M samples
	Input data matrix size	48097	18
	After sample removal	48097	18
	After features with NA>40% removal	23663	18
	After filtering out nonproteotypic peptides	21604	18
	After peptides summarization	3589	18
	Final protein matrix size	3589	18
Cerebellum			
	Input data matrix size	39908	21
	After sample removal	39908	21
	After features with NA>40% removal	28074	21
	After filtering out nonproteotypic peptides	25736	21
	After peptides summarization	4008	21
	Final protein matrix size	4008	21
Brainstem			
	Input data matrix size	50699	24
	After sample removal	50699	22
	After features with NA>40% removal	28236	22
	After filtering out nonproteotypic peptides	25994	22
	After peptides summarization	3857	22
	Final protein matrix size	3857	22
Mucosa			
	Input data matrix size	39966	22
	After sample removal	39966	21
	After features with NA>40% removal	17962	21
	After filtering out nonproteotypic peptides	16426	21
	After peptides summarization	2969	21
	Final protein matrix size	2969	21

Final distribution of samples in each clinical group is presented in **Table 3**

Table 3. Distribution of samples across clinical groups after quality control.

Class\tissue	Bulbus	Cerebellum	Medulla	Mucosa	Total in Class
acute	6	8	8	8	30
late	7	9	9	8	33
control	5	4	5	5	19
Total in tissue	18	21	22	21	82

Statistical analysis

Statistical analysis of proteomics data was carried out using internally developed R scripts. Linear modelling was based on the R package LIMMA[33]. Following model was applied to each tissue data set (log(p) is log2 transformed expression of a protein): log(p) ~ 0+Class.

The categorical factor Class has three levels: control, acute, late; Reference level: late. The following contrasts were evaluated in each tissue type: Contrast1: acute – control; Contrast2: late – control; Contrast3: late – acute.

For finding regulated features following criteria were applied: Significance level alpha was set to guarantee on average false discovery rate below ~ 30% across all contrasts in the tissue and across all tissues in the contrast. We found that alpha = 0.01 was delivering the required level of Benjamini–Hochberg FDR[34] below 30% on average, see **Table 4**. Going for a more stringent significance level (lower alpha or FDR threshold) was considered unreasonable, since we wanted to compare different contrasts within tissue and between tissues.

Table 4. Maximum level of Benjamini-Hochberg FDR obtained in regulated features in each tissue\contrast with significance level a = 0.01. Overall average FDR value is 0.3. Contrast1: acute – control; Contrast2: late – control; Contrast3: late – acute.

Tissue\Contrast	Contrast1	Contrast2	Contrast3	Tissue Avg
Bulbus	0.38	0.5	0.28	0.39
Cerebellum	0.36	0.26	0.21	0.28
Brainstem	0.29	0.4	0.22	0.30
Mucosa	0.13	0.17	0.36	0.22
Contrast Avg	0.29	0.33	0.27	0.30

The log fold change criterion was applied to guarantee that the measured signal is above the average noise level. As such we have taken median residual standard deviation of linear model: log2(T) = median residual SD of linear modelling, see **Table 5** for specific values.

Table 5. Fold change threshold for each tissue type.

Tissue	Bulbus	Cerebellum	Brainstem	Mucosa
Fold Change Threshold T	1.37	1.23	1.25	1.45

Functional analysis of proteomics data

Functional analysis was carried out using R packages gprofiler2[35] for overrepresentation analysis and clusterProfiler[36] for GSEA. For selecting the most (de)regulated GO terms we applied filter: 2 ≤ term size ≤ 250. Unless specified separately, analyses were carried out with Benjamini-Hochberg FDR threshold 0.3. For ORA, statistical domain scope was set to custom, list of quantified proteins was used as background.

Materials Proteomics

Hydrophobic Sera-Mag magnetic carboxylate modified particles (44152105050250 Fisher Scientific), hydrophilic Sera-Mag magnetic carboxylate modified particles (24152105050250 Fisher Scientific), Twintec skirted low bind plates (0030129512 Eppendorf), TCEP (646547 Sigma Aldrich), SDS (A7249.1000 Applichem), CAA (22788 Merck/Millipore), ammonium bicarbonat (/871.2 Roth), 100 % ACN (955-212 Fisher Scientific), 80 % ethanol (1.00983.2500 Millipore), 230 µl Biomek Tips (B85903 Beckmann Coulter), Eppendorf 500 µl deep well plates (30501101 Eppendorf), Waters Acquity UPLC 700 µl plates 990 (186005837 Waters GmbH) Sequencing grade modified Trypsin (V5117 Promega), Pierce Quantitative Fluorometric Peptide Assay (number 23290), formic acid (85178 Thermo Scientific), water (1.15333.2500 Merck), protease inhibitor cocktail complete mini (Roche 994 04693124001), benzonase nuclease (Sigma Aldrich E1014-25KU).

Supplementary References

1. Gassen, N.C.*, et al.* SARS-CoV-2-mediated dysregulation of metabolism and autophagy uncovers host-targeting antivirals. *Nat Commun* **12**, 3818 (2021).
2. Young, M.D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**(2020).
3. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
4. Stuart, T.*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
5. Brann, D.H.*, et al.* Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Sci Adv* **6**(2020).
6. Durante, M.A.*, et al.* Single-cell analysis of olfactory neurogenesis and differentiation in adult humans. *Nat Neurosci* **23**, 323-326 (2020).
7. Khrameeva, E.*, et al.* Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* **30**, 776-789 (2020).
8. Darmanis, S.*, et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* **112**, 7285-7290 (2015).
9. Chua, R.L.*, et al.* COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol* **38**, 970-979 (2020).

10. Bottcher, C.*, et al.* Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. *Nat Neurosci* **22**, 78-90 (2019).
11. Winkler, E.A.*, et al.* A single-cell atlas of the normal and malformed human brain vasculature. *Science* **375**, eabi7377 (2022).
12. Faure, L.*, et al.* Single cell RNA sequencing identifies early diversity of sensory neurons forming via bi-potential intermediates. *Nat Commun* **11**, 4175 (2020).
13. Polioudakis, D.*, et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785-801 e788 (2019).
14. Synaptic Neurotransmission: GABAergic Inhibition. Vol. 2022 Version 5.12.701.700 (Weizmann Institute of Science, 2022).
15. Nicotine Pathway (Dpaminergic Neuron), Pharmacodynamics. Vol. 2022 Version 5.12.701.700 (Weizmann Institute of Science, 2022).
16. Lab., L. Mouse Brain Atlas. Vol. 2022.
17. The Jackson Laboratory, B.H., Maine. The Mouse Genome Database (MGD), Mouse Genome Informatics. Vol. 2022.
18. Blake, J.A.*, et al.* Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res* **49**, D981-D987 (2021).
19. Smith, C.M.*, et al.* The mouse Gene Expression Database (GXD): 2019 update. *Nucleic acids research* **47**, D774-D779 (2019).
20. Buttner, M., Ostner, J., Muller, C.L., Theis, F.J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun* **12**, 6876 (2021).
21. Holzwarth, K.*, et al.* Multiplexed fluorescence microscopy reveals heterogeneity among stromal cells in mouse bone marrow sections. *Cytometry A* **93**, 876-888 (2018).
22. Pascual-Reguant, A.*, et al.* Multiplexed histology analyses for the phenotypic and spatial characterization of human innate lymphoid cells. *Nat Commun* **12**, 1737 (2021).
23. Schubert, W.*, et al.* Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nature Biotechnology* **24**, 1270-1278 (2006).
24. Pertuz, S., Puig, D., Garcia, M.A. & Fusiello, A. Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images. *IEEE Trans Image Process* **22**, 1242-1251 (2013).
25. Schindelin, J.*, et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).
26. Rusinova, I.*, et al.* Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res* **41**, D1040-1046 (2013).

27. Hao, Y.*, et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).
28. Koopmans, F.*, et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234 e214 (2019).
29. Josse, J. & Husson, F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software* **70**, 1 - 31 (2016).
30. Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).
31. Cox, J.*, et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526 (2014).
32. Pham, T.V., Henneman, A.A. & Jimenez, C.R. iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. *Bioinformatics* **36**, 2611-2613 (2020).
33. Ritchie, M.E.*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
34. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
35. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**(2020).
36. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).