Brain imaging before and after COVID-19 in UK Biobank

Gwenaëlle Douaud[1✉], Soojin Lee[1], Fidel Alfaro-Almagro[1], Christoph Arthofer[1], Chaoyue Wang[1], Paul McCarthy[1], Frederik Lange[1], Jesper L.R. Andersson[1], Ludovica Griffanti[1,2], Eugene Duff[1,3], Saad Jbabdi[1], Bernd Taschler[1], Anderson M. Winkler[4], Thomas E. Nichols[5], Rory Collins[6], Paul M. Matthews[7], Naomi Allen[6], Karla L. Miller[1], Stephen M. Smith[1]

[1]*FMRIB Centre, Wellcome Centre for Integrative Neuroimaging (WIN), Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK*

[2]*OHBA, Wellcome Centre for Integrative Neuroimaging (WIN), Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK*

[3]*Department of Paediatrics, University of Oxford, Oxford, UK*

[4]*National Institutes of Mental Health, National Institutes of Health, Bethesda, MD, USA*

[5]*Big Data Institute, University of Oxford, Oxford, UK*

[6]*Nuffield Department of Population Health, University of Oxford, Oxford, UK*

[7]*UK Dementia Research Institute and Department of Brain Sciences, Imperial College, London, UK*

✉ Correspondence to: Prof. Gwenaëlle Douaud

FMRIB Centre, Wellcome Centre for Integrative Neuroimaging,

Nuffield Department of Clinical Neurosciences,

John Radcliffe Hospital, Oxford OX3 9DU, UK

+44 1865 610 478

gwenaelle.douaud@ndcn.ox.ac.uk

Abstract

There is strong evidence for brain-related pathologies in COVID-19, some of which could be a consequence of viral neurotropism, or of neuroinflammation following viral infection. Most brain imaging studies have focused on qualitative, gross pathology in moderate to severe cases, most typically carried out on hospitalised patients. It remains unknown however whether the impact of SARS-CoV-2 infection can be detected in milder cases, in a quantitative and automated manner, and whether this can reveal possible mechanisms for the spread of the disease. UK Biobank scanned over 40,000 participants before the start of the COVID-19 pandemic, making it possible in 2021 to invite back hundreds of previously-imaged participants for a second imaging visit. Here, we studied the possible brain changes associated with the coronavirus infection using multimodal MRI data from 785 adult participants (aged 51–81) from the UK Biobank COVID-19 re-imaging study, including 401 adult participants who tested positive for SARS-CoV-2 infection between their two scans. We used structural, diffusion and functional brain scans from before and after infection, to compare longitudinal changes between these 401 SARS-CoV-2 cases and 384 controls who had either tested negative to rapid antibody testing or had no COVID-19 medical and public health record, and who were matched to the cases for age, sex, ethnicity and interval between scans. The controls and cases did not differ in blood pressure, body mass index, diabetes diagnosis, smoking, alcohol consumption, or socio-economic status. Using both hypothesis-driven and exploratory approaches, with false discovery rate multiple comparison correction, we identified respectively 68 and 67 significant longitudinal effects associated with SARS-CoV-2 infection in the brain, including, on average: (i) a more pronounced reduction in grey matter thickness and contrast in the lateral orbitofrontal cortex (min $P=1.7\times10^{-4}$, r=-0.14) and parahippocampal gyrus (min $P=2.7\times10^{-4}$, r=-0.13), (ii) a relative increase of diffusion indices, a marker of tissue damage, in the regions of the brain functionally-connected to the piriform cortex, anterior olfactory nucleus and olfactory tubercle (min $P=2.2\times10^{-5}$, r=0.16), and (iii) greater reduction in global measures of brain size and increase in cerebrospinal fluid volume suggesting an additional diffuse atrophy in the infected participants (min $P=4.0\times10^{-6}$, r=-0.17). When looking over the entire cortical surface, these grey matter thickness results covered the parahippocampal gyrus and the lateral orbitofrontal cortex, and extended to the anterior insula and anterior cingulate cortex, supramarginal gyrus and temporal pole. The increase of

a diffusion index (mean diffusivity) meanwhile could be seen voxel-wise mainly in the medial and lateral orbitofrontal cortex, the anterior insula, the anterior cingulate cortex and the amygdala. These results were not altered after excluding cases who had been hospitalised. We further compared hospitalised (n=15) and non-hospitalised (n=386) infected participants, resulting in similar findings to the larger cases vs control group comparison, with, in addition, a marked reduction of grey matter thickness in fronto-parietal and temporal regions (all FDR-significant, min $P=4.0\times10^{-6}$). The 401 SARS-CoV-2 infected participants also showed larger cognitive decline between the two timepoints in the Trail Making Test compared with the controls (both FDR-significant, min $P=1.0\times10^{-4}$, r=0.17; and still FDR-significant after excluding the hospitalised patients: min $P=1.0\times10^{-4}$, r=0.17), with the duration taken to complete the alphanumeric trail correlating *post hoc* with the cognitive and olfactory-related crus II of the cerebellum (FDR-significant, $P=2.0\times10^{-3}$, r=-0.19), which was also found significantly atrophic in the SARS-CoV-2 participants (FDR-significant, $P=6.1\times10^{-5}$, r=-0.14). Our findings thus relate to longitudinal abnormalities in limbic cortical areas with direct neuronal connectivity to the primary olfactory system. Unlike in *post hoc* cross-sectional studies, the availability of pre-infection imaging data mitigates to some extent the issue of pre-existing risk factors or clinical conditions being misinterpreted as disease effects. We were therefore able to demonstrate that the regions of the brain that showed longitudinal differences post-infection did not already show any difference between (future) cases and controls in their initial, pre-infection scans. These brain imaging results may be the *in vivo* hallmarks of a degenerative spread of the disease — or of the virus itself — via olfactory pathways (a possible entry point of the virus to the central nervous system being via the olfactory mucosa), or of neuroinflammatory events due to the infection, or of the loss of sensory input due to anosmia. Whether this deleterious impact can be partially reversed, for instance after improvement of the hyposmic symptoms, or whether these are effects that will persist in the long term, remains to be investigated with additional follow up.

Introduction

While the global pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has now claimed millions of lives across the world, there has been increased focus by the scientific and medical community on the effects of mild-to-moderate COVID-19 in the longer term. There is strong evidence for brain-related pathologies, some of which could be a consequence of viral neurotropism[1,2], or of virus-induced neuroinflammation[3-5](doi.org/10.1101/2021.02.23.432474): neurological and cognitive deficits demonstrated by patients[6,7], with an incidence of neurological symptoms in more than 80% of the severe cases[8], radiological and *post mortem* tissue analyses demonstrating the impact of COVID-19 on the brain[9,10], and the possible presence of the coronavirus in the central nervous system found in some studies[11-13].

In particular, one consistent clinical feature, which can appear before the onset of respiratory symptoms, is the disturbance in olfaction and gustation in COVID-19 patients[14,15]. In a recent study, 100% of the patients in the subacute stage of the disease were displaying signs of gustatory impairment (hypogeusia), and 86% either hyposmia or anosmia[16]. Such loss of sensory olfactory inputs to the brain could lead to a loss of grey matter in olfactory-related brain regions[17]. Olfactory — whether neuronal or supporting — cells concentrated in the olfactory epithelium are also particularly vulnerable to coronavirus invasion, and this seems to be also the case specifically with SARS-CoV-2[15,18-20]. Within the olfactory system, direct neuronal connections from and to the olfactory bulb encompass regions of the piriform cortex (the primary olfactory cortex), parahippocampal gyrus, entorhinal cortex, and orbitofrontal areas[21,22].

Most brain imaging studies of COVID-19 to date have focused on acute cases and radiological reports of single cases or case series based on CT, PET or MRI scans, revealing a broad array of gross cerebral abnormalities ranging from white matter hyperintensities, hypoperfusion and signs of ischaemic events spread throughout the brain, but found more consistently in the cerebrum[9]. Of the few larger studies focusing on cerebrovascular damage using CT or MRI, some have either found no clear marker of abnormalities in the majority of their patients, or importantly no spatially consistent pattern for the distribution of white matter hyperintensities or microhaemorrhages, except perhaps in the middle or posterior cerebral artery territories and the basal ganglia[9]. Imaging cohort studies of COVID-19, quantitatively

comparing data across subjects through automated preprocessing and co-alignment of images, are much rarer. For instance, a recent PET cohort study focusing on correlates of cognitive impairment has demonstrated, in 29 COVID-19 patients at a subacute stage, the involvement of fronto-parietal areas revealed as ^{18}F-FDG hypometabolism[16]. Another glucose PET study has shown bilateral hypometabolism in the bilateral orbital gyrus rectus and the right medial temporal lobe[23]. One multi-organ imaging study[24] (and its brain-focused follow-up - doi.org/10.1101/2021.05.19.21257316) in over 50 previously hospitalised COVID-19 patients suggested modest abnormalities in T2* of the left and right thalami compared with matched controls. It remains unknown however whether any of these abnormalities *predates* the infection by SARS-CoV-2. These effects could be associated with a pre-existing increased brain vulnerability to the deleterious effects of COVID-19 and/or a higher probability to show more pronounced symptoms, rather than being a consequence of the COVID-19 disease process.

UK Biobank offers a unique resource to elucidate these questions. With the data from this large, multi-modal brain imaging study, we use for the first time a longitudinal design whereby participants had been already scanned as part of UK Biobank *before* getting infected by SARS-CoV-2. They were then imaged again, on average 38 months later, after some had either medical and public health records for COVID-19, or had had two positive rapid antibody tests. Those participants were then matched with controls who had undergone the same longitudinal imaging protocol but had tested negative to the rapid antibody test or had no medical record of COVID-19. In total, 401 SARS-CoV-2 infected participants with usable imaging data at both timepoints were included in this study, as well as 384 controls, matched for age, sex, ethnicity and time elapsed between the two scans. These large numbers may allow us to detect subtle, but consistent spatially distributed sites of damage caused by the infection, thus underlining *in vivo* the possible spreading pathways of the effects of the disease within the brain (whether such effects relate to the invasion of the virus itself[11,18], inflammatory reactions[3,4] doi.org/10.1101/2021.02.23.432474, possible anterograde degeneration starting with the olfactory neurons in the nose, or through sensory deprivation[17,25,26]). The longitudinal aspect of the study should in turn make it possible to tease apart which of the observed effects between first and second scans are likely related to the infection.

Our general approach in this study was therefore as follows: (i) use brain imaging data from 785 participants who visited the UK Biobank imaging centres for two scanning sessions, on average 3 years apart, with 401 of these having been infected with SARS-CoV-2 in between their two scans; (ii) estimate — from each subject's multimodal brain imaging data — hundreds of distinct brain imaging-derived phenotypes (IDPs), each IDP being a measure of one aspect of brain structure or function; (iii) model confounding effects, and estimate the longitudinal change in IDPs between the two scans; and (iv) identify significant SARS-CoV-2 vs control group differences in these longitudinal effects, correcting for multiple comparisons across IDPs. We did this for both a focussed set of *a priori* defined IDPs, testing the hypothesis that the olfactory system is particularly vulnerable in COVID-19, as well as an exploratory set of analyses considering a much larger set of IDPs. In both cases we identified significant effects associated with SARS-CoV-2 infection primarily relating to greater atrophy and increased tissue damage in cortical areas directly connected to primary olfactory cortex, as well as to changes in global measures of brain and cerebrospinal fluid volume.

Methods

Study Design

As part of the UK Biobank imaging study[27], thousands of subjects had received brain scans before the start of the COVID-19 pandemic. Multimodal brain imaging data, collected at four sites with identical imaging hardware, scanner software and protocols, and passing quality controls, was obtained from 42,729 participants over the age of 45 years, and made available to researchers worldwide.

Before the COVID-19 pandemic, longitudinal (first- and second-timepoint scanning) had already begun in the UK Biobank imaging study, with up to 3,000 participants returning for a second scan prior to scanning being paused in 2020 as a result of the pandemic. More recently, starting in February 2021, hundreds of UK Biobank participants who had already taken part in UK Biobank imaging before the pandemic were invited back for a second scan. This COVID-19 re-imaging study was set up to investigate the effects of SARS-CoV-2 infection by comparing imaging scans taken from participants before vs after infection.

The full list of inclusion criteria for the participants in this re-imaging study is as follows:
- had already attended an imaging assessment at one of the three imaging sites (the fourth opened just before the pandemic began),
- still lived within the catchment area of the clinic they attended for their first imaging assessment,
- had no incidental findings identified from their scans taken at the first imaging visit,
- had not withdrawn or died
- had a valid email and postal address,
- had high-quality scans from the first imaging visit,
- lived within 60 km of the clinic (extended to 75 km in Feb 2021), due to travel restrictions during the lockdown period.

(See for more details the online documentation: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/casecontrol_covidimaging.pdf)

Amongst those, some participants were identified as having been infected with SARS-CoV-2 based on: (i) their primary care (GP) data or hospital records, or (ii) results of diagnostic antigen tests identified through linkage to health-related records, or (iii) results of two

antibody tests. Indeed, participants were also invited to take a home-based lateral flow (Fortress Fast COVID-19 Home test, Fortress Diagnostics and ABC-19™ Rapid Test, Abingdon Health) to detect the presence of SARS-CoV-2 antibodies. A second kit was sent to all participants who recorded an initial positive result and who had indicated they had not yet been vaccinated, in order to reduce the number of false positives.

Controls were then selected by identifying, from the remaining previously imaged UK Biobank participants, those who had a negative antibody test result, as determined from the home-based lateral flow kits, and/or who had no record of confirmed or suspected COVID-19 from primary care, hospital records or diagnostic antigen test data. Controls were selected to match 1:1 to positive SARS-CoV-2 cases according to five criteria:

- sex
- ethnicity (white/non-white, as numbers were too low to allow for a finer distinction)
- date of birth (+/-6 months)
- location of first imaging assessment clinic
- date of first imaging assessment (+/-6 months).

Image Processing

For this work, we primarily used the IDPs generated by our team on behalf of UK Biobank, and made available to all researchers by UK Biobank[27,28]. The IDPs are summary measures, each describing a different aspect of brain structure or function, depending on what underlying imaging modality is used[27,28].

The protocol includes three structural MRI scans (T1, T2 fluid attenuation inversion recovery (FLAIR) and susceptibility-weighted MRI), as well as diffusion MRI and resting and task functional MRI. T1 scans make it possible to derive global measures of brain and cerebrospinal fluid (CSF) volumes, as well as localised measures of grey matter volume and cortical thickness and area. The T2 FLAIR scan identifies differences that might be indicative of inflammation or tissue damage. Susceptibility-weighted MRI is sensitive to iron and myelin content. Diffusion MRI measurements give insight into the tissue microstructure integrity. Resting-state functional MRI is performed on an individual who is not engaged in any particular activity or

task, and can provide indices related to the functional connectivity between brain regions[29]. Functional connectivity is intrinsically noisy when each region-pair connection is considered individually, so we focused here our analysis on 6 dimensionally-reduced functional connectivity networks[30].

We used 1,524 existing UK Biobank IDPs, including: regional grey matter, brain and CSF volume, local cortical surface area, volume and thickness, cortical grey-white contrast, white matter hyperintensity volume, white matter microstructural measures such as fractional anisotropy and mean diffusivity, resting-state amplitude and dimensionally-reduced connectivity measures. In addition, we also generated 1,106 new IDPs, as described below.

We computed additional IDPs obtained using Quantitative Susceptibility Mapping (QSM), which has been recently added into our UK Biobank processing pipeline (doi.org/10.1101/2021.06.28.450248). Magnitude and phase data from the susceptibility-weighted MRI acquisitions were processed to provide quantitative measures reflecting clinically relevant tissue susceptibility properties. Median T2* was calculated within 17 subcortical structures (with their regions-of-interest (ROIs) estimated from the T1) as IDPs; 14 of these are the same subcortical regions already estimated by the core UK Biobank pipeline, and here we added 3 more subcortical ROIs: left and right substantia nigra[31] and regions of white matter hyperintensities (lesions)[32]. Second, susceptibility-weighted MRI phase data were processed for QSM following a pipeline recently developed for UK Biobank[33] (doi.org/10.1101/2021.05.19.21257316). QSM (CSF-referenced) IDPs were calculated in the same 17 subcortical structures as the T2* IDPs.

Additional IDPs were created via refined sub-segmentations of the hippocampus, amygdala and thalamus as implemented in FreeSurfer[34-37]. We extracted these ROI masks from the FreeSurfer processing and applied them to the T2* and diffusion images (diffusion tensor model: MD and FA; NODDI model: OD, ISOVF, ICVF) to generate additional subcortical IDPs.

Finally, we generated new IDPs tailored to the olfactory and gustatory systems, as described below.

Olfactory and gustatory systems hypothesis-driven approach

Based on prior expectations from animal models and *post mortem* findings, we chose to focus *a priori* our primary analyses on a subset of 332 regions-of-interest (297 of which passed the reproducibility thresholding; see Reproducibility section below) from the available 2,630 IDPs[21,22,38]; these correspond anatomically to the telencephalic primary and secondary connections of the olfactory and gustatory cortex. Briefly, these include the piriform cortex, parahippocampal gyrus, entorhinal cortex, amygdala, insula, frontal/parietal operculum, medial and lateral orbitofrontal cortex, hippocampus and basal ganglia. As no labelling of the piriform cortex exists in any of the atlases used in the UK Biobank imaging processing, we refined a previously published ROI of the piriform cortex (frontal and temporal), anterior olfactory nucleus and olfactory tubercle, by limiting it to the cortical ribbon of our UK Biobank T1-weighted standard space (https://github.com/zelanolab/primaryolfactorycortex parcellation[39]). We further used maps from the same study's resting-state fMRI analysis of the functional connectivity of each of the four parts of this ROI (piriform frontal, piriform temporal, anterior olfactory nucleus and olfactory tubercle) to the rest of the brain, to generate four additional extended ROIs of the functionally-connected cortical and subcortical regions to these primary olfactory areas[39]. For this, we thresholded their connectivity t-value maps to keep only significant voxels ($P_{fwe}<0.05$, with threshold-free cluster enhancement), and used the maps as weighted (and, separately, binarised) masks, to further extract grey matter volume (GM), T2* and diffusion values; this was done by: (i) regressing each of these maps into the GM, T2* or diffusion images in their respective native spaces, and separately, (ii) by binarising the maps and extracting mean and 95th percentile values.

Additionally, masks for the left and right olfactory bulbs were generated by manually drawing a binary mask for the right olfactory bulb on an averaged template-space T2 FLAIR volume generated from 713 UK Biobank subjects, and mirroring this to obtain the mask for the left (having confirmed by visual inspection that symmetry in this region allowed for this to be effective). Both masks were then modulated by the T2 intensities in their respective ROIs, to account for partial volume effects, generating the final "label" maps with values ranging between 0-1. For the hypothalamus, we combined and refined ROIs from two previously published and publicly available atlases of a probabilistic hypothalamus map (https://neurovault.org/collections/3145/[31]) and hypothalamic subregions[40]. Both the probabilistic hypothalamus map and the

binarised map obtained from fusing the 26 hypothalamic subregions were transformed to our standard space where the probabilistic map was then masked by the binarised map. We then extracted volume, and T2 mean and 95th percentile intensity measurements in subjects' native spaces, using the olfactory bulb and hypothalamus maps (unthresholded and thresholded at 0.3). For the hypothalamus, we additionally extracted these metrics from T2* and diffusion images. All of the above preprocessing steps were defined and completed before any analyses of longitudinal change and case-control modelling.

The full list of 297 pre-determined and reproducible IDPs is available in **Table S1.**

Exploratory Approach

The full set of 2,630 IDPs described above were used for a more exploratory, inclusive analysis of SARS-CoV-2 infection effects on brain structure and function (see full list of reproducible IDPs in **Table S2**).

Statistical Modelling

The following modelling was applied in the same way to both the hypothesis-driven analyses of a subset of IDPs, and the all-IDPs exploratory analyses.

Automated outlier identification of the IDPs

All IDPs from all subjects were pooled for initial processing (at this stage blinded to the SARS-CoV-2 status of participants): 42,729 Scan 1 datasets (all pre-pandemic), 2,943 pre-pandemic Scan 2 datasets, and 890 Scan 2 datasets acquired after the beginning of the COVID-19 pandemic. Extreme outlier values (individual IDPs from individual scanning sessions) were removed on the basis of being more extreme than 8 times the median absolute deviation from the median for a given IDP.

The IDPs from the 890 subjects imaged during the pandemic (SARS-CoV-2 positive cases and controls), from both timepoints, were then retained. Subjects were kept if at least the T1-weighted structural image was usable from both timepoints, resulting in IDPs at both timepoints (IDP1 and IDP2) from 785 subjects. The data were then pooled into a single dataset comprising 785 × 2 = 1,570 imaging sessions, and cross-sectional deconfounding, treating all

scans equivalently, was carried out for head size, scanner table position, and image motion in the diffusion MRI data (these imaging confound variables first had outlier removal applied as described above, though using a higher threshold of 15 times the median absolute deviation, given the highly non-Gaussian nature of some confounds)[28,41].

Reproducibility of the IDPs

We then evaluated the scan-rescan reproducibility of IDPs, in order to discard IDPs that were not reasonably reproducible between scans. For each IDP, we correlated the IDP1 with IDP2 values, separately for cases and controls, resulting in two reproducibility measures (Pearson correlation r) for each IDP. The (N_{IDPs} long) vectors of r values derived from cases and from controls were extremely highly correlated (r=0.98), showing that potential effects associated with infection are subtle compared with between-subject variability and IDP noise; hence, we averaged these cases and controls' r values to give a single reproducibility measure for each IDP. From the initial set of 2,630 IDPs, the least reproducible IDPs (r<0.5) were discarded, leaving 2,048 IDPs. Finally, IDPs with high levels of missing data (usable values from fewer than 50 subjects) were discarded, leaving in total 2,047 IDPs.

Deconfounding and main longitudinal model

Despite initial case-control subject pairing (resulting in case and control groups being well matched), missing/outlier data potentially disrupted this exact paired matching, and thus we also included in the modelling confound variables derived from those factors originally used as pairing criteria: difference between the subjects' ages at each of their two scans, the difference of the squares of the ages (to account for quadratic dependencies of IDPs on age), genetic sex, and ethnicity (white vs non-white).

Longitudinal IDP change (ΔIDP) was estimated by regressing IDP2 on IDP1[42], as well as the confound variables listed above. The case-control difference in this longitudinal IDP effect was modelled with a group difference regressor comprised of the case-vs-control binary variable modulated by a function of age at Scan 2 (Age2, that being a close proxy for age at infection for the SARS-CoV-2 group, with less than a year's error); as the effects of COVID-19 increase exponentially with age[43,44], we used the exact age dependence found by a meta-regression of

28 studies to modulate the binary case-vs-control variable, based on age at Scan 2[44]. The main group difference regressor of interest is therefore:

$$\text{Case-vs-Control} = \text{demeaned(Case-vs-Control binary variable)} \times 10^{\text{Age2}\times 0.0524-3.27}, \quad [1]$$

where these age-dependence parameters are taken from the meta-regression analysis[44]. To ensure that the fitting of this term is not influenced by an effect that is common to controls and cases, we added an "Age2-Exponential-Effect" confound variable of $10^{\text{Age2}\times 0.0524-3.27}$, i.e., the same aging term without the group-difference multiplier.

Our main model of interest is therefore simply:

$$\text{IDP2} \sim \text{Case-vs-Control} + \text{IDP1} + \text{Confounds}, \quad [2]$$

where the Confounds comprise the terms described above: Age2-Age1, Age2^2-Age1^2, ethnicity, sex, and Age2-Exponential-Effect.

Instead of fitting the regression in Eq (2), we residualise the response and the group difference regressor with respect to the nuisance (IDP1 and confounds) and then compute their Pearson correlation to find the strength of the association between the longitudinal IDP effect and the age-modulated case-vs-control effect. This correlation approach gives identical inferences to regression but is a more convenient form when using permutation testing to derive exact family-wise error corrected P values. The correlations were Fisher-transformed to Z-statistics, with the standard scaling by sqrt(N-3), to take into account the different patterns of missing data for different IDPs. For each IDP, any missing data is ignored (that subject is left out for that analysis). The Z-statistics are valuable because they provide a standardised measure of effect size, accounting for varying degrees of freedom in different IDPs. As part of the estimation of the longitudinal IDP changes, ΔIDP outliers (for each IDP, and each subject) were removed (set as missing), if they were more than 8 times the median absolute deviation from the median.

Permutation testing and correction for multiple comparisons

We used permutation testing to estimate family-wise-error P-values (P_{fwe}), i.e., correcting for the multiple comparisons across IDPs while accounting for the dependences among IDPs. We

randomly permuted the residualised case-vs-control regressor relative to the residualised IDP2s, with 10,000 random permutations. At each permutation we computed the association Z value for each IDP, and recorded the maximum absolute value across all IDPs. By taking the absolute value, we corrected for the two-tailed nature of the test, i.e., we did not pre-assume the direction of any effect. After building up the null distribution of the maximum |Z| across IDPs, we then tested the original |Z| values against this distribution, to obtain family-wise error corrected P-values (P_{fwe}), fully correcting for multiple comparisons across all IDPs. We also computed for each test the false discovery rate (FDR) at 5%, generating a threshold that can be applied to uncorrected P-values to determine their FDR significance.

Group comparisons

In the rest of the manuscript, we refer to the main age-modulated-group comparison analysis (comparing IDPs at second timepoint controlling for IDPs at baseline) between SARS-CoV-2 positive cases and controls, as described above, as Model 1. As secondary analyses, we also applied the same hypothesis-driven and exploratory approaches as described above, to compare non-hospitalised SARS-CoV-2 positive cases against controls (Model 2), and hospitalised patients against controls (Model 3). Separately, we also carried out the same analysis between hospitalised and non-hospitalised cases, adding as covariates three risk factors showing significant differences between these two SARS-CoV-2 groups (Model 4). For these secondary models (2-4), we again used age-modulated group-difference regressors as described above for Model 1. Power to detect effects in the two latter models, considering the hospitalised patients as a separate group, is of course considerably reduced, given the small number of hospitalised cases in this cohort.

For all 4 models, testing was carried out twice: first using the *a priori* focussed subset of IDPs identified for the hypothesis-driven analyses, and then using the full set of IDPs for the exploratory analyses. In both cases IDPs were identified as having significant group differences, corrected for multiple comparisons.

Cognitive analysis

While cognitive testing offers limited measurements of cognitive function in UK Biobank, we included 9 variables in our analysis related to tests that commonly detect neurological abnormalities[45-47]:

- three variables from the Trail Making Test: both durations to complete trails A and B, as well as the total number of errors made traversing trail B, based on the most discriminating measures of cognitive impairment from this test based on prior literature[45],

- two variables based on Matrix Reasoning ("matrix pattern recognition"): the number of puzzles correctly solved, and number of puzzles viewed,

- one variable from the Symbol Digit Test: the number of symbol digit matches made correctly,

- one measure of reaction time: mean time to correctly identify matches at the card game "Snap",

- one measure of reasoning: the "fluid intelligence" score,

- one measure of delayed recall: the result of the prospective memory test.

For these 9 variables from 6 different cognitive tests, we carried out two analyses: (i) the same cases vs controls group comparison, of the longitudinal effect in cognitive scores (ΔCOG), as described above for testing changes in IDPs, with the case-vs-control modulated by age regressor as described above, (ii) a *post hoc* correlation analysis, in the SARS-CoV-2 group only, of the ΔCOG showing the most significant difference between cases and controls against the top most significant ΔIDP across hypothesis-driven and exploratory approaches. All results were evaluated for FWE and FDR significance, correcting for multiple comparisons across all cognitive, or IDP variables where applicable.

Additional analyses

Baseline cross-sectional group comparisons

<u>Risk factors</u>: We compared the SARS-CoV-2 positive and control groups at baseline (modelled using a simple binary regressor) across common risk factors for infection and severity of disease: age, sex, blood pressure (systolic and diastolic), weight (including BMI, and waist-hip

ratio), diabetes, smoking, alcohol consumption and socio-economic status (via the Townsend deprivation index). For this, we used the 'Last Observation Carried Forward' (LOCF) imputation method, for which we considered all the values available closest to the Scan 1 visit (for the majority of the values, these were available from the same visit, on the same day that Scan 1 was acquired); we also tested that there was no difference between SARS-CoV-2 and control groups in the distribution of the visits used to collect the LOCF values.

All other non-imaging phenotypes: We also examined whether the two SARS-CoV-2 and control groups differed at baseline across all non-imaging phenotypes (lifestyle, environmental, health-related, dietary), across all UK Biobank visits. We assessed the 7,545 non-imaging phenotypes having at least 3% of values as being distinct from the majority value, and results were corrected for multiple comparisons using FDR and FWE (i.e., where relevant we refer to both in Results).

IDPs: To complement our longitudinal analyses, we carried out a baseline-only (and, separately, second timepoint only) cross-sectional group comparison between SARS-CoV-2 cases and controls (modelled using a binary regressor) across all 2,047 IDPs, correcting for multiple comparisons across all IDPs using the same permutation-testing procedure as described above.

In particular, this approach is of interest to test whether brain regions showing significant longitudinal changes demonstrate initial differences, pre-existing *before* the infection, between the two groups.

Cognition: We finally assessed whether the two groups differed at baseline in their cognition, based on the results from the 9 variables from 6 different cognitive tests preselected above, correcting for multiple comparisons across cognitive variables.

Lateralised effects

As a *post hoc* analysis, we explored whether the longitudinal effects observed in grey matter thickness was lateralised, by subtracting right ΔIDP from the corresponding left ΔIDP, for: (i) all ΔIDPs of grey matter thickness showing significance in the main case-control analyses (across the hypothesis-driven and exploratory approaches), within the SARS-CoV-2 group only

(to avoid circularity); (ii) all ΔIDPs of grey matter thickness across the entire cortex (151 pairs of left-right matched IDPs), and correlating the left-right difference against the case-vs-control age modulated regressor. Results were corrected for multiple comparisons using FDR and FWE.

Effect of time between SARS-CoV-2 infection and brain scan post-infection

For 351 SARS-CoV-2 positive participants who had a date available for infection (hence, in effect excluding those identified through antibody lateral flow tests), we further looked *post hoc* at the possible effect of time interval between infection and second brain scan (acquired post-infection) on the significant IDP from our hypothesis-driven approach, to evaluate whether a longer interval might mean either a reduced loss of grey matter through potential progressive recovery of sensory inputs (olfaction), or greater loss as a function of a longer, ongoing degenerative process.

Post hoc assessment of potential impact of non-imaging factors on longitudinal IDP effects

We ran an additional analysis to test whether any non-imaging variables measured before SARS-CoV-2 infection might explain the longitudinal effects observed in our significant IDPs. We considered non-imaging variables with at least 80% non-missing data in the participants (n=1,734). We regressed these out of the ΔIDPs and the main case-vs-control regressor, including them as additional confounds for a repeat of the original Model 1 regression tests for those IDPs found to show significant longitudinal differences between the two groups, for both hypothesis-driven and exploratory approaches. If the strength of the original association was reduced by more than 50%, based on the correlation coefficient (as r is not biased by changes in degrees of freedom due to missing non-imaging data), we considered a non-imaging variable to potentially explain the IDP-infection association.

Results

Participants

UK Biobank has been releasing data from the COVID-19 re-imaging study on a rolling basis. As of the 31st of May 2021, 449 adult participants met the re-imaging study inclusion criteria (see **Methods**) and were identified as having been infected with SARS-CoV-2 based on either their primary care (GP) data, hospital records, results of their diagnostic antigen tests identified through record linkage to the Public Health datasets in England, Wales and Scotland, or two concordant antibody-based home lateral flow kit positive results. Of these 449 SARS-CoV-2 positive adult participants, a total of 401 had usable brain scans at both timepoints (**Tables 1 and 2**). In total, 384 adult controls met the inclusion criteria (see **Methods**) and had usable brain scans at both timepoints (**Table 1**). SARS-CoV-2 positive or negative status was identified using UK Biobank Showcase variable 41000.

Despite the original matched-pairing of the COVID-19 patients and controls, their age distributions were slightly — though not statistically significantly — different, due to different patterns of missing/usable data (**Figure S1**). Note that the control group is on average slightly (not significantly) older than the SARS-CoV-2 positive group, which would be expected to make any change between the two timepoints harder to detect in the group comparisons, rather than easier.

For histograms of interscan intervals in the two groups, see **Figure S2**.

The two groups showed no statistical differences across all 7,545 non-imaging phenotypes after FDR or FWE correction for multiple comparisons (lowest P_{fwe}=0.12, and no uncorrected P values survived FDR correction).

Table 1. Main demographics for the 401 SARS-CoV-2 positive cases and 384 controls. We used the 'Last Observation Carried Forward' (LOCF) imputation method (**Methods**). Non-parametric tests were used whenever a variable for each group was not normally distributed (Lilliefors $P < 0.05$). Two-sample Kolmogorov-Smirnov test was used for age at Scan 1 or Scan 2, years between Scan 1 and Scan 2, alcohol intake frequency, and tobacco smoking; chi-square test for sex, ethnicity, and diagnosed diabetes; and Mann-Whitney U-test was used for the systolic and diastolic blood pressures, weight, waist-hip ratio, BMI and Townsend deprivation index.

	SARS-CoV-2 Positive Cases	Controls	P_{uncorr}
Number of subjects	401	384	-
Age at Scan 1, mean±SD (range)	58.9 ± 7.0 (46.9–80.2)	60.2 ± 7.4 (47.1–79.8)	0.15
Age at Scan 2, mean±SD (range)	62.1 ± 6.7 (51.3–81.4)	63.3 ± 7.1 (51.3–81.3)	0.08
Sex, male/female	172 (42.9%) / 229 (57.1%)	164 (42.7%) / 220 (57.3%)	0.96
Ethnicity, white/non-white*	388 (96.8%) / 13 (3.2%)	373 (97.1%) / 11 (2.9%)	0.76
Years between Scans 1 and 2, mean±SD (range)	3.2 ± 1.6 (1.0–7.0)	3.2 ± 1.6 (1.0–6.9)	0.98
Systolic blood pressure [mmHg]	130.3 ± 17.3	132.1 ± 17.6	0.16
Diastolic blood pressure [mmHg]	78.7 ± 10.6	79.0 ± 10.2	0.63
Diagnosed diabetes	18 (4.5%)	16 (4.2%)	0.82
Weight [kg]	76.4 ± 15.8	75.2 ± 14.4	0.65
Waist/Hip ratio	0.87 ± 0.09	0.86 ± 0.09	0.37
BMI [kg/m^2]	26.7 ± 4.4	26.6 ± 4.3	0.61
Alcohol intake frequency	3.1 ± 1.3	3.0 ± 1.4	1.00
Tobacco smoking	0.61 ± 0.92	0.65 ± 0.89	0.87
Townsend deprivation index	-1.5 ± 2.9	-1.6 ± 2.9	0.65

*The white/non-white distinction was made as numbers were too low to allow for a finer distinction

Table 2. Main clinical information available for the SARS-CoV-2 positive cases. Of note, of the 401 participants in our SARS-CoV-2 positive group in our main analyses, 50 were identified as cases via two different antibody home-based lateral flow kits and do not have date of diagnosis in their primary care or hospital records.

	N, or mean ± SD (range)
Total number of positive cases	**401**
Origin of diagnosis	
- GP	11
- Hospital	2
- Diagnostic antigen test from Public Health records	338
- Antibody home-based lateral flow kits	50
Number of infected participants with available information on date of diagnoses	**351**
Date of diagnosis	
- 01/01/2020 – 31/03/2020	4
- 01/04/2020 – 30/06/2020	29
- 01/07/2020 – 30/09/2020	18
- 01/10/2020 – 31/12/2020	200
- 01/01/2021 – 31/03/2021	99
- 01/04/2021 – 31/05/2021	1
ΔDays of SARS-CoV-2 infection before Scan 2, mean±SD (range)	141±79 (35–407)
Total number of hospitalised patients	**15**
- COVID-19 as primary cause	11
- COVID-19 as secondary cause	4

-	Days of hospitalization, mean±SD (range)	11.1 ±11.0 (1–40)
-	Critical care unit	2
-	Invasive ventilation	1
-	Continuous positive airway pressure	1
-	Non-invasive ventilation	1
-	Unspecified oxygen therapy	1

Through hospital records available for participants, we identified 15 of the SARS-CoV-2 positive group who were hospitalised with COVID-19, including 2 who received critical care (**Tables 2 and 3**). These hospitalised patients were on average older, had higher blood pressure and weight, and were more likely to have diabetes and to be men, compared with non-hospitalised cases (**Table 3**).

Table 3. Comparison between hospitalised vs non-hospitalised SARS-CoV-2 positive cases. For statistical procedures, please refer to **Table 1**.

	Hospitalised	**Non-hospitalised**	P_{uncorr}
Number of subjects	15	386	-
Age at Scan 1, mean±SD (range)	65.4 ± 8.9 (51.6–80.2)	58.7 ± 6.8 (46.9–77.0)	0.0028
Age at Scan 2, mean±SD (range)	68.1 ± 8.4 (54.9–81.4)	61.9 ± 6.5 (51.3–80.0)	0.0058
Sex, male/female	10 (66.7%) / 5 (33.3%)	162 (42.0%) / 224 (58.0%)	0.058
Ethnicity, white/non-white*	15 (100%) / 0 (0%)	373 (96.6%) / 13 (3.4%)	0.47
Years between Scan 1 and 2, mean±SD (range)	2.7 ± 1.4 (1.0–5.8)	3.2 ± 1.6 (1.1–7.0)	0.50
Systolic blood pressure [mmHg]	140.6 ± 16.6	129.9 ± 17.2	0.022
Diastolic blood pressure [mmHg]	85.0 ± 10.5	78.4 ± 10.5	0.028
Diagnosed diabetes	4 (26.7%)	14 (3.6%)	< 0.001
Weight [kg]	85.9 ± 12.0	76.0 ± 15.8	0.0072
Waist/Hip ratio	0.94 ± 0.07	0.87 ± 0.09	0.0015
BMI [kg/m^2]	29.3 ± 3.7	26.6 ± 4.4	0.0076
Alcohol intake frequency	3.1 ± 1.7	3.1 ± 1.3	1.00
Tobacco smoking	0.80 ± 1.0	0.60 ± 0.91	0.75
Townsend deprivation index	-2.1 ± 2.6	-1.5 ± 2.9	0.42

*The white/non-white distinction was made as numbers were too low to allow for a finer distinction

Hypothesis-driven results

The case-vs-control analysis between the 401 SARS-CoV-2 positive cases and 384 controls (Model 1) on 297 olfactory-related cerebral IDPs yielded 68 significant results after FDR correction for multiple comparisons, including 6 further surviving FWE correction (**Table 4**, **Figure 1, Table S1** for full list of results). Focusing on the top 10 most significant associations, 8 of these IDPs covered similar brain regions functionally-connected to the primary olfactory cortex (see Methods), showing overlap especially in the anterior cingulate cortex, orbito-frontal cortex and insula, as well as in the ventral striatum, amygdala, hippocampus and parahippocampal gyrus[39]. We found greater longitudinal increase in diffusion indices for the SARS-CoV-2 group in these tailored IDPs defining the functional connections with the frontal and temporal piriform cortex, as well as the olfactory tubercle and anterior olfactory nucleus (**Table 4**, **Figure 1, Table S1**). The other two of the top 10 IDPs encompassed the left lateral orbitofrontal cortex and parahippocampal gyrus, both showing greater reduction of grey matter thickness or intensity contrast over time in the cases compared with controls (**Table 4**, **Figure 1, Table S1**).

Significant longitudinal differences remained in the same set of significant brain regions surviving FDR or FWE correction when removing from the SARS-CoV-2 group those patients who had been hospitalised with COVID-19 (Model 2, 47 IDPs FDR-significant, 3 of which also FWE-significant, **Table S1**).

While fewer results were significant for the comparison between the 15 hospitalised patients and 384 controls (Model 3, 4 results FDR-significant, **Table S1**), likely due to the large reduction in sample size for this model, this additional group comparison showed effects in the same regions of the parahippocampal gyrus, orbital cortex, and superior insula.

We found no significant differences between the 15 hospitalised patients and 386 non-hospitalised SARS-COV-2 cases, likely due to the large reduction in sample size, but effect sizes suggested similar effects once more in the orbitofrontal, insula, parahippocampal and frontal piriform cortex functionally-connected brain regions (all $|Z|\geq 3$, Model 4, **Table S1**).

Across the 3 models comparing SARS-CoV-2 cases with controls (Models 1-3), the top 4 longitudinal differences were found in the functionally-connected regions of the temporal piriform cortex (diffusion index: orientation dispersion) and of the olfactory tubercle

(diffusion index: isotropic volume fraction), as well as in the parahippocampal gyrus (intensity contrast) and lateral orbitofrontal cortex (thickness) (largest combined |Z| across Models 1-3; **Figure 1**). For these results across Models 1-3, the percentage of SARS-CoV-2 infected participants who showed a greater longitudinal change than the median value in the controls was: 56% for the regions connected to the temporal piriform cortex, 62% for the regions connected to the olfactory tubercle, 57% left parahippocampal gyrus and 60% for the left orbitofrontal cortex.

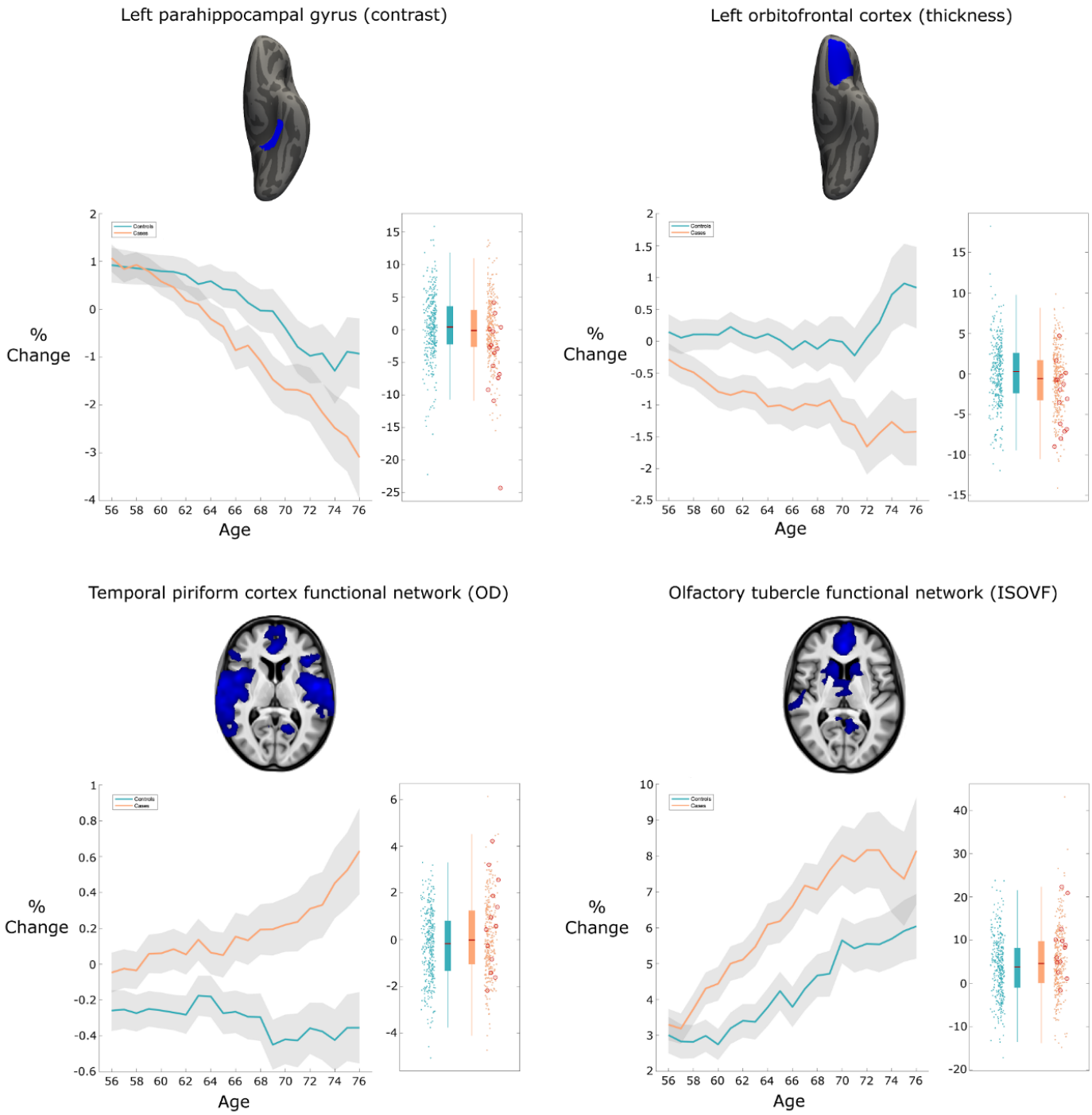

Figure 1. Most significant longitudinal group comparison results - hypothesis-driven approach. The top four regions consistently showing longitudinal differences across the three models comparing SARS-CoV-2 cases and controls demonstrated either a significantly greater reduction in grey matter thickness and intensity contrast, or an increase in tissue damage (largest combined |Z| across Models 1-3). All three models pointed at the involvement of the parahippocampal gyrus, while Models 1 and 2 also showed the significant involvement of the functional connections of the primary olfactory cortex and of the left orbitofrontal cortex. For each region, the IDP's spatial region of interest is shown at the top in blue, overlaid either on the FreeSurfer average inflated cortical surface, or the T1 template (left is shown on right). Bottom left for each IDP are the longitudinal percentage changes with age for the two groups (controls in blue, infected participants in orange), obtained by normalising ΔIDP using as baseline the values for the corresponding IDPs across up to 46,743 scans from the larger UK Biobank imaging study. These are created using a 10 year sliding window average, with standard errors in grey. The somewhat counter-intuitive increase in thickness in the orbitofrontal cortex in older controls has been previously consistently reported in studies of ageing[48]. Bottom right are the scatter and box plots showing the difference in cortical thickness, contrast intensity, or diffusion indices between the two timepoints for the 384 controls (blue) and 401 infected participants (orange), allowing the visual comparison between the two groups in a binary way (hence under-estimating the effects estimated when modulating with age, see **Methods**). In red circles are the 15 hospitalised patients. OD, orientation dispersion; ISOVF, isotropic volume fraction. All y axes represent % change.

While significant IDPs related to grey matter thickness were found, using Model 1, to be bilateral for both the anterior parahippocampal gyrus (perirhinal cortex) and entorhinal cortex, 10 of the 11 remaining significant IDP were left-lateralised (**Table S1**). We thus directly investigated (left - right) differences in the SARS-CoV-2 group only for those significant IDPs, and found that the infected participants did not have significantly more reduced grey matter thickness on the left than on the right hemisphere (lowest P_{uncorr}=0.30).

None of the top 10 most significant results showed a significant effect of interval between infection and second scan in the SARS-CoV-2 positive participants for whom we had a date of diagnosis (n=351; lowest P_{uncorr}=0.08).

Exploratory results

2,047 IDPs passed the initial tests of reproducibility (**Figure S3**) and data completeness. The main analysis (Model 1) revealed 67 significant longitudinal differences between the cases and controls passing FDR correction, including 5 that were FWE-significant (**Table 5**, **Table S2** for the complete list of reproducible IDPs and results). **Figures S4** and **S5** show the QQ plot relating to the FDR thresholding, and a summary figure of Z-statistics results for all 2,047 IDPs grouped into different IDP classes.

In particular, in this exploratory analysis covering the entire brain, 32 out of the 67 significant IDPs overlapped with the IDPs selected *a priori* for our hypothesis-driven approach of the involvement of the olfactory system. In addition, we found significant longitudinal effects in global measures of volume, such as the CSF volume normalised for head size and the ratio of the volume of the segmented brain to the estimated total intracranial volume generated by FreeSurfer, as well as in the volume of the left crus II of the cerebellum, the thickness of the left rostral anterior cingulate cortex and diffusion index in the superior fronto-occipital fasciculus (**Table 5**, **Table S2**, see examples in **Figure S6**).

When comparing the non-hospitalised cases to the controls (Model 2), the same general pattern emerged, albeit with a reduced number of significant results: one olfactory-related region, the functionally-connected areas to the temporal piriform cortex, showed significant longitudinal difference between the two groups in diffusion index, as well as one global volume measure (CSF normalised), and diffusion index in the superior fronto-occipital fasciculus (Model 2, 4 FDR-corrected, 1 FWE-corrected, **Table S2**).

Despite the considerably limited degrees of freedom in Models 3 and 4, many results survived multiple comparison correction, particularly for IDPs of cortical thickness, with an emphasis on the anterior cingulate cortex for Model 3 (85 FDR-corrected, 3 FWE-corrected), and a wide distribution across prefrontal, parietal and temporal lobes for Model 4 (47 FDR-corrected, 1 FWE-corrected).

As many of the top exploratory and hypothesis-driven results included IDPs of cortical thickness and of mean diffusivity, we further conducted an exploratory visualisation of the vertex-wise longitudinal thickness, and voxel-wise longitudinal mean diffusivity differences between the cases and controls over the entire cortical surface and brain volume, respectively

(**Figure 2**). Cortical grey matter thickness showed bilateral longitudinal differences in the parahippocampal gyrus, anterior cingulate cortex and temporal pole, as well as in the left orbitofrontal cortex, insula and supramarginal gyrus. When comparing hospitalised and non-hospitalised cases, these longitudinal differences showed a similar pattern, especially in the parahippocampal gyrus, orbitofrontal and anterior cingulate cortex, but also markedly extending, particularly in the left hemisphere, to many fronto-parietal and temporal regions. Mean diffusivity differences in longitudinal effects between cases and controls was seen mainly in the orbitofrontal cortex, anterior cingulate cortex, as well as in the left insula and amygdala.

While results seen in IDPs of grey matter thickness seemed to indicate that the left hemisphere is more strongly associated with SARS-CoV-2 infection, a direct (left - right) comparisons of all lateralised IDPs of thickness across the entire cortex showed no overall statistical difference between the two groups (lowest P_{fwe}=0.43, and with no results surviving FDR correction).

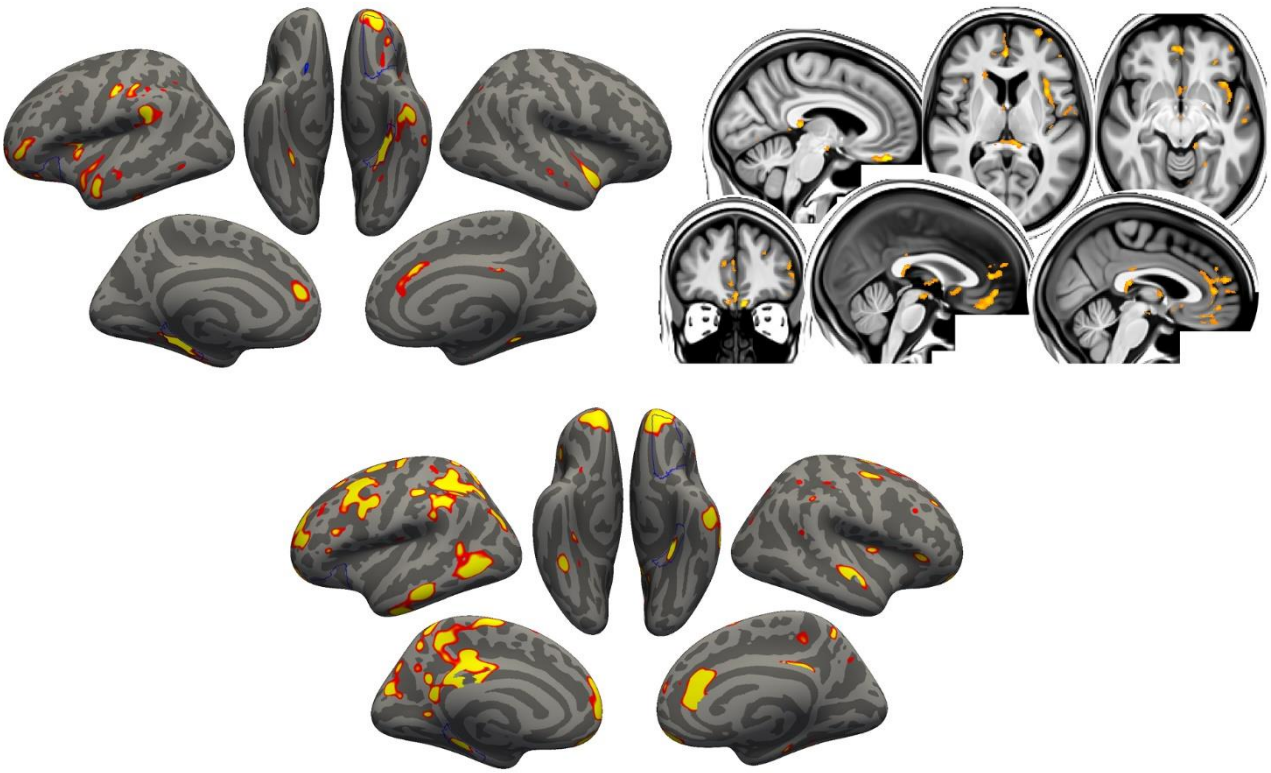

Figure 2. Vertex-wise and voxel-wise longitudinal group comparison results in grey matter thickness and mean diffusivity. Top row. Model 1: The thresholded map (Z>3) shows that the strongest, localised reduction of grey matter thickness in the 401 SARS-CoV-2 positive participants compared with the 384 controls are bilaterally in the parahippocampal gyrus, anterior cingulate cortex and temporal pole, as well as in the left orbitofrontal cortex, insula and supramarginal gyrus. Similarly, the strongest longitudinal differences in mean diffusivity (Z>3, left is shown on right) could be seen in the orbitofrontal cortex and anterior cingulate cortex, as well as in the left insula and amygdala. **Bottom row.** Model 4: The thresholded cortical thickness map (Z>3) demonstrated longitudinal differences between the 15 hospitalised and 386 non-hospitalised SARS-CoV-2 positive cases in the orbitofrontal frontal cortex and parahippocampal gyrus bilaterally, right anterior cingulate cortex, as well as marked widespread differences in fronto-parietal and temporal areas, especially in the left hemisphere.

Table 4. Hypothesis-driven olfactory approach: longitudinal group comparison results. The top 10 significant results, all surviving false discovery rate (FDR) correction, out of 297 imaging-derived phenotypes (IDPs), ranked based on their uncorrected P-values for Model 1, showing where the 401 SARS-CoV-2 infected participants and 384 controls differed over time. Associations with a total of 68 IDPs survived correction for multiple comparisons using FDR for Model 1 (full list of results in **Table S1**). All significant results involved either grey matter thickness, grey-white intensity contrast or proxy measures of tissue damage (mean diffusivity MD, isotropic volume fraction ISOVF, and orientation dispersion OD). Results in bold also survive correction for multiple comparisons using family-wise error (FWE) for each corresponding Model. Results in italics survive correction for multiple comparisons using FDR for each corresponding Model. The Z-statistics reflect the statistical strength of the longitudinal group-difference modelling, and are not raw data effect sizes. L is left.

	Model 1: All SARS-CoV-2 cases (n=401) vs controls (n=384)				**Model 2: SARS-CoV-2 non-hospitalised cases (n=386) vs controls (n=384)**			**Model 3: SARS-CoV-2 hospitalised cases (n=15) vs controls (n=384)**			**Model 4: SARS-CoV-2 hospitalised (n=15) vs non-hospitalised cases (n=386)**		
Imaging-Derived Phenotype (IDP) Description	**r**	**Z**	**Puncorr**	**Pfwe**	**Z**	**Puncorr**	**Pfwe**	**Z**	**Puncorr**	**Pfwe**	**Z**	**Puncorr**	**Pfwe**
Temporal piriform cortex functional network - OD	***0.16***	***4.3***	***0.000022***	***0.0065***	***4.0***	***0.000077***	***0.0190***	2.5	0.013334	0.9169	0.5	0.623737	1
Olfactory tubercle functional network - ISOVF	***0.15***	***3.9***	***0.000096***	***0.0243***	*3.4*	*0.000595*	*0.1250*	2.9	0.004171	0.6356	1.4	0.150892	1
Frontal piriform cortex functional network - MD (mean)	***0.14***	***3.8***	***0.000138***	***0.0343***	*3.4*	*0.000641*	*0.1347*	2.4	0.016953	0.9505	1.3	0.197097	1
Temporal piriform cortex functional network - MD	***0.14***	***3.8***	***0.000142***	***0.0352***	*3.4*	*0.000813*	*0.1662*	2.5	0.011175	0.8833	1.4	0.167831	1
Olfactory tubercle functional network - MD	***0.14***	***3.8***	***0.000162***	***0.0400***	*3.3*	*0.000906*	*0.1835*	2.5	0.01128	0.885	1.4	0.168789	1
Lateral orbitofrontal cortex L - thickness (DKT atlas)	***-0.14***	***-3.8***	***0.000165***	***0.0408***	*-3.0*	*0.003078*	*0.4670*	-3.1	0.001898	0.4301	-1.9	0.059449	0.9999
Temporal piriform cortex functional network - ISOVF	*0.14*	*3.7*	*0.000211*	*0.0512*	*3.3*	*0.001024*	*0.2031*	2.6	0.009338	0.8461	1.3	0.178746	1
Anterior olfactory nucleus functional network - MD	*0.14*	*3.7*	*0.00023*	*0.0547*	*3.5*	*0.000534*	*0.1127*	1.9	0.058551	0.9997	0.8	0.427243	1
Parahippocampal gyrus L - contrast intensity (Desikan atlas)	*-0.13*	*-3.7*	*0.000265*	*0.0635*	*-2.7*	*0.006033*	*0.6854*	*-3.5*	*0.000488*	*0.1909*	-2	0.040896	0.9986
Anterior olfactory nucleus functional network - ISOVF	*0.14*	*3.6*	*0.000272*	*0.0651*	*3.3*	*0.000944*	*0.1906*	2.4	0.015347	0.9387	1.1	0.263033	1

Table 5. Exploratory approach: longitudinal group comparison results. The top 10 significant results show where the 401 SARS-CoV-2 positive participants and 384 controls differed over time, ranked based on their uncorrected P-values for Model 1. Associations with a total of 67 imaging-derived phenotypes (IDPs) survived correction for multiple comparisons using false discovery rate (FDR) for Model 1 (full list of results in **Table S2**). In bold, the findings surviving correction for multiple comparisons using family-wise error (FWE) for each Model; in italics, those surviving FDR. In addition to global measures relating to loss of brain volume (such as an increase of CSF volume), most of the top exploratory localised results implicate the primary connections of the olfactory system, as well as the rostral anterior cingulate cortex and the crus II of the cerebellum, both also olfactory-related regions. The Z-statistics reflect the statistical strength of the longitudinal group-difference modelling, and are not raw data effect sizes. Intra-cellular volume fraction ICVF, orientation dispersion OD. L is left, R is right.

	Model 1: All SARS-CoV-2 cases (n=401) vs controls (n=384)				**Model 2: SARS-CoV-2 non-hospitalised cases (n=386) vs controls (n=384)**			**Model 3: SARS-CoV-2 hospitalised cases (n=15) vs controls (n=384)**			**Model 4: SARS-CoV-2 hospitalised (n=15) vs non-hospitalised cases (n=386)**		
Imaging-Derived Phenotype (IDP) Description	**r**	**Z**	**Puncorr**	**Pfwe**	**Z**	**Puncorr**	**Pfwe**	**Z**	**Puncorr**	**Pfwe**	**Z**	**Puncorr**	**Pfwe**
Ratio brain volume/estimated total intracranial volume	***-0.17***	***-4.6***	***0.000004***	***0.0079***	-3.3	0.001137	0.7795	*-4.6*	*0.000005*	*0.0511*	*-3.4*	*0.000709*	*0.7138*
Normalised CSF - volume	***0.15***	***4.3***	***0.000015***	***0.0279***	*4.1*	*0.000045*	*0.0750*	1.8	0.067173	1	0.5	0.616664	1
Lateral ventricle R - volume	***0.16***	***4.3***	***0.000019***	***0.0351***	3.7	0.000234	0.3055	2.7	0.006719	0.9985	1.2	0.215132	1
Temporal piriform cortex functional network - OD	***0.16***	***4.3***	***0.000022***	***0.040***	*4.0*	*0.000077*	*0.1203*	2.5	0.013334	0.9999	0.5	0.623737	1
Superior fronto-occipital fasciculus - ICVF	***-0.16***	***-4.2***	***0.000024***	***0.0421***	***-4.3***	***0.000016***	***0.0319***	-1.1	0.27516	1	0.6	0.580107	1
Brain volume without ventricles (surface model estimate)	*-0.15*	*-4.1*	*0.000041*	*0.0691*	-3.0	0.003186	0.9766	*-4.0*	*0.000067*	*0.248*	-2.7	0.006771	0.9994
Rostral anterior cingulate cortex L - thickness (Desikan atlas)	*-0.15*	*-4.1*	*0.000042*	*0.0701*	-2.9	0.003712	0.9870	*-4.3*	*0.000019*	*0.1175*	-2.3	0.020912	1
Brain volume without ventricles	*-0.15*	*-4.1*	*0.000043*	*0.0731*	-2.9	0.003329	0.9801	*-4.0*	*0.00007*	*0.2563*	-2.7	0.006612	0.9993
Supratentorial volume without ventricles	*-0.15*	*-4.0*	*0.000055*	*0.0902*	-2.9	0.003262	0.9787	*-3.8*	*0.00016*	*0.3990*	-2.5	0.011406	1
Cerebellum crus II - volume	*-0.14*	*-4.0*	*0.000061*	*0.0985*	-3.1	0.001919	0.9107	*-3.3*	*0.000863*	*0.7900*	-2.1	0.035186	1

Cognitive results

Using the same model used to compare longitudinal effects in IDPs between SARS-CoV-2 positive participants and controls (Model 1), we explored differences between the two groups in 9 scores from 6 cognitive tasks. After Bonferroni correction for (these 9) multiple comparisons, we found a significantly greater increase of the time taken to complete Trails A (numeric) and B (alphanumeric) of the Trail Making Test in the SARS-CoV-2 infected group (Trail A: r=0.16, P_{fwe}=0.003; Trail B: r=0.17, P_{fwe}=0.0009; **Figure 3**).

In the SARS-CoV-2 group only, *post hoc* correlations between the most significant cognitive score showing longitudinal effect using Model 1 (duration to complete Trail B, as reported above) and the top 10 results from each of the hypothesis-driven and exploratory approaches revealed a significant longitudinal association with the volume of the mainly cognitive lobule crus II of the cerebellum (r=-0.19, P_{fwe}=0.04, Bonferroni-corrected across the 20 IDPs).

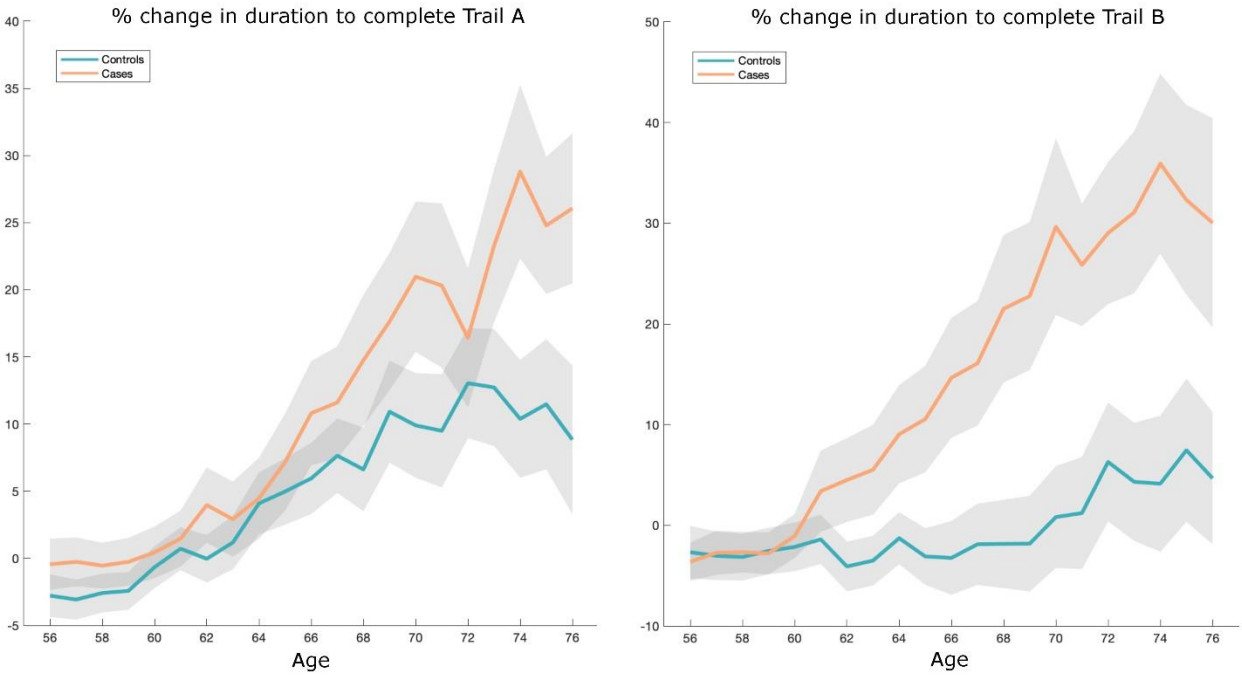

Figure 3. Percentage longitudinal change for SARS-CoV-2 positive participants and controls, in the duration to complete Trails A and B of the Trail Making Test. Absolute baseline (used to convert longitudinal change into percent change) estimated across the 785 participants. These curves were created using a 10 year sliding window across cases and controls (standard errors in grey).

Additional cross-sectional results

None of the IDPs with significant longitudinal effects for either hypothesis-driven or exploratory approaches demonstrated significant differences at baseline between SARS-CoV-2 and control groups (lowest P_{fwe}=1.00, nothing surviving FDR correction; **Table S3**). The full list of results from group comparisons between the two groups at baseline, and separately, at the second timepoint are available in **Tables S3 and S4**.

In addition, none of the 9 cognitive variables showed significant difference at baseline between SARS-CoV-2 and control groups (min P_{uncorr}=0.08).

Additional tests of potential influence of baseline non-imaging factors on significant results

We repeated Model 1 association tests for those IDPs found to show longitudinal differences between the SARS-CoV-2 and control groups, across both hypothesis-driven and exploratory approaches. For each of 1,734 non-imaging variables available (see **Methods**), we included that variable as an additional confounder in the longitudinal analyses. On the basis of the correlation coefficient, the strength of the original associations was not reduced by more than 50% for any of the non-imaging variables.

Discussion

To our knowledge, this is the first longitudinal imaging study of SARS-CoV-2 where participants were initially scanned *before* any had been infected. This multimodal exploration of the longitudinal effects in the brain associated with SARS-CoV-2 infection was made possible by leveraging the unique resource of the UK Biobank project, and its COVID-19 re-imaging sub-study. 785 participants had two usable brain scans: 401 SARS-CoV-2 positive participants (cases), and 384 controls who were well-matched for age, sex, ethnic background and interval between the two scans, as well as various risk factors. The hypothesis-driven and exploratory longitudinal analyses revealed a significant, deleterious impact associated with SARS-CoV-2. This impact could be seen mainly in the olfactory cortical system, for instance with a change in diffusion measures — that are proxies for tissue damage — in regions functionally connected with the piriform cortex, olfactory tubercle and anterior olfactory nucleus, as well as a more pronounced reduction of grey matter thickness and contrast in the SARS-CoV-2 infected participants in the left parahippocampal gyrus and lateral orbitofrontal cortex. While the greater atrophy for the SARS-CoV-2 positive participants was localised to a few, mainly limbic, regions, the increase in CSF volume and decrease of whole brain volume suggests in addition a diffuse loss of grey matter superimposed onto the more regional effects observed in the olfactory-related areas, something that seems further confirmed by the shift in the histograms of grey matter thickness Z-statistics (**Figure S7**). Further investigation demonstrated that this effect associated with infection, both in the cortical thickness and mean diffusivity, extended mainly to the anterior cingulate cortex, insula, supramarginal gyrus and temporal pole. Comparing the few patients (n=15) who had been hospitalised with COVID-19 against non-hospitalised cases showed a more widespread pattern of greater reduction in grey matter thickness in fronto-parietal and temporal regions (**Figure 2**). Finally, significantly greater cognitive decline, which persisted even after excluding the hospitalised patients, was seen in the SARS-CoV-2 positive group between the two timepoints, and this decline was associated with greater atrophy of crus II, a cognitive lobule of the cerebellum.

Much has been made of the benefit of using a longitudinal design to estimate (for example) trajectories of brain ageing and cognitive decline[49] (doi.org/10.1101/2021.02.08.428915). The longitudinal nature of the UK Biobank COVID-19 re-imaging study, with the baseline scan acquired *before* infection by SARS-CoV-2 and the second scan *after* infection, makes it

possible to more confidently disentangle the contribution of the pathogenic process from pre-existing differences in the brain (or risk factors) of future COVID-19 patients. None of the significant longitudinal differences between the two groups were also reflected as a significant cross-sectional group difference at baseline (lowest P_{fwe}=1.00); that is, before infection, the participants in each group did not exhibit significant limbic, olfactory-related differences. Another illustrative example of the benefit of a longitudinal design is that, if looking solely at cross-sectional group comparisons at the second timepoint post infection (i.e., the analysis that would, by necessity, be carried out in *post hoc* studies), the strongest effect is seen in the volume of the thalamus. This effect disappears when taking into account the baseline scans however, since the thalamus of the participants who will later become infected appears to already differ from the controls years before infection. Whether these differences played a subsequent role in those patients being more likely to get infected by the coronavirus, or to develop symptoms from them, would need to be investigated further by linking them to either genetic information, or the lifestyle, cognitive, physical and other health outcome measures available in UK Biobank. This highlights the difficulties in interpreting cross-sectional post-infection differences as being necessarily the consequence of the infection itself. The longitudinal aspect of this study design, combined with the matching of the patients and controls for age, sex, scan interval and ethnicity, increases the interpretability of the results, and confidence that the longitudinal abnormalities observed in those olfactory-related brain regions are an impact of the infection. As this study is observational (as opposed to a randomised interventional study), one cannot make claims of disease causality with absolute certainty, but interpretational ambiguities are greatly reduced compared with *post hoc* cross-sectional studies.

Our cohort-based, quantitative imaging study, unlike the majority of single case and case series studies published so far, does not focus on gross abnormalities that could be observed at the single subject-level with a naked eye, such as microhaemorrhages or (sub)acute ischaemic infarctions[9]. It does however rely on an anatomically *consistent* pattern of abnormalities caused by the disease process, a common spatial distribution of these pathological alterations across the infected participants, which could be uncovered by aligning all the images together in a common space, followed by applying a pipeline of modality-specific image processing algorithms. This automated, objective and quantitative

processing of the images facilitates the detection of subtle changes that would not be visible at the individual level, but which point to a possible mechanism for the neurological effects of the coronavirus infection. Our hypothesis-driven analyses revealed a clear involvement of the olfactory cortex, which was also found in the exploratory analyses and the vertex-wise and voxel-wise maps of cortical thickness and mean diffusivity. While no differences were seen in the olfactory bulbs or piriform cortex *per se* (both located in a region above the sinuses prone to susceptibility distortions in the brain images, and both being difficult to segment in MRI data), we identified significant longitudinal differences in a network of regions functionally-connected to the piriform cortex, mainly constituted of the anterior cingulate cortex and orbitofrontal cortex, as well as the ventral striatum, amygdala, hippocampus and parahippocampal gyrus[39]. Some of the most consistent abnormalities across hypothesis-driven and exploratory analyses and all group comparisons were revealed in the left parahippocampal gyrus (**Table 4**, **Figure 2**, **Tables S1 and S2**) — a limbic region of the brain that plays a crucial, integrative role for the relative temporal order of events in episodic memory[50,51,52]. Importantly, it is directly connected to the piriform cortex and entorhinal cortex, which are both part of the primary olfactory cortex[22,53]. Similarly, the orbitofrontal cortex, which we also found altered in the SARS-CoV-2 positive group, is often referred to as the secondary olfactory cortex, as it possesses direct connections to both entorhinal and piriform cortex[53], as well as to the anterior olfactory nucleus[21,39]. In fact, in a recent functional connectivity study of the primary olfactory cortex, the orbitofrontal cortex was found to be connected to all four primary olfactory regions investigated (frontal and temporal piriform cortex, anterior olfactory nucleus and olfactory tubercle), possibly explaining why it is reliably activated even in basic and passive olfactory tasks[39]. Using the same olfactory connectivity maps, which overlap cortically in the orbitofrontal cortex, anterior cingulate cortex and insula, we found a more pronounced increase of diffusion metrics indicative of tissue damage in the SARS-CoV-2 group. The voxel-wise map of mean diffusivity pinpointed these longitudinal differences in the orbitofrontal and anterior cingulate cortex, as well as in the insula and the amygdala. The insula is not only directly connected to the primary olfactory cortex[21], but is also considered to be the primary gustatory cortex. “Area G” (i.e., the dorsal part of the insula at the junction with the frontal and parietal operculum), in turn connects with the orbitofrontal cortex[38]. The vertex-wise and voxel-wise visualisation of both greater loss of grey matter and increase in mean diffusivity in the insula spatially correspond in particular to

the area of consistent activation to all basic taste qualities[54]. Finally, the exploratory analysis revealed a more pronounced loss of grey matter in crus II, part of the cognitive, and olfactory-related lobule VII of the cerebellum[55]. These results are in line with previous post-infection PET findings showing, in more severe cases, FDG hypometabolism in the insula, orbitofrontal and anterior cingulate cortex, as well as lower grey matter volume in the insula and hippocampus[56,57].

Early neurological signs in COVID-19 include hyposmia and hypogeusia, which appear to precede the onset of respiratory symptoms in the majority of affected patients[2,18,58]. In addition, a heavily-debated hypothesis has been that an entry point of SARS-CoV-2 to the central nervous system is via the olfactory mucosa, or the olfactory bulb[2,11,18]. (The coronavirus itself would not necessarily need to enter the central nervous system; anterograde degeneration from olfactory neurons might suffice to generate the pattern of abnormalities revealed in our longitudinal analyses.) The predominance observed in other studies of hyposmic and anosmic symptoms — whether caused directly by loss of olfactory neurons or by perturbation of supporting cells of the olfactory epithelium[15,20] — could also, through repeated sensory deprivation, lead to loss of grey matter in these olfactory-related brain regions. Very focal reduction in grey matter in the orbitofrontal cortex and insula have been observed for instance in patients with severe olfactory dysfunction in a cross-sectional study of chronic rhinosinusitis[26]. A more extensive study of congenital and acquired (post-infectious, chronic inflammation due to rhinosinusitis, or idiopathic) olfactory loss also demonstrated an association between grey matter volume and olfactory function in the orbitofrontal cortex[17]. It also showed that duration of olfactory loss for those with acquired olfactory dysfunction, ranging from 0 to over 10 years, was related to more pronounced loss of grey matter in the gyrus rectus and orbitofrontal cortex[17]. On the other hand, it has been reported in a longitudinal study that patients with idiopathic olfactory loss had higher grey matter volume after undergoing olfactory training in various brain regions including the orbitofrontal cortex and gyrus rectus[59]. This raises the interesting possibility that the pattern of longitudinal abnormalities observed here in the limbic, olfactory brain regions of SARS-CoV-2 positive participants, if they are indeed related to olfactory dysfunction, might be attenuated over time if the infected participants go on to recover their sense of smell and taste. There is for instance some very preliminary evidence, in a few previously hospitalised

COVID-19 patients, that brain hypometabolism becomes less pronounced when followed-up 6 months later, even if it does not entirely resolve[56,60]. When we tested whether time between infection and second brain scan had any relationship — positive, indicative of recovery, or negative, indicative of an ongoing degenerative process — with the grey matter loss or increase in diffusivity in the significant IDPs, we found no significant effect, possibly owing to the relatively small range in duration of infection at the time of this study (between 1 and 13 months for those 351 infected participants for whom we had a diagnosis date), and particularly with only a few having been infected for over 6 months. Additional follow-up of this cohort would be particularly valuable in determining the longer-term effects of infection on these limbic structures.

An alternative explanation for the brain differences revealed in this longitudinal study might be that the coronavirus infection leads to neuroinflammation, which can initiate chronic neuronal dysfunction[3,4]. In particular, activation of the peripheral innate immune system can induce the production of inflammatory cytokines in the brain, leading not only to severe impairment in memory, cognition and emotional behaviour, but also to microglial abnormalities in the hippocampus, which is particularly vulnerable to neuroinflammation[61-63]. In line with these findings, in the case of an averred neurotropic virus such as influenza, which spreads to central nervous system via infection of the olfactory neurons in the olfactory epithelium[64], various long-term inflammatory-induced functional and structural alterations of the hippocampus, accompanied by impairment in spatial memory, have been recently observed[65]. It is also possible that the choroid plexus, which appears to relay peripheral inflammation due to SARS-CoV-2 into the brain[3], plays a role in the specific targeting of olfactory regions, as they might regulate the dynamics of the subventricular zone responsible for the neurogenesis occurring in the olfactory bulbs[66]. (Maladaptive) immune regulation through the microglia might also modulate these neurogenic niches[67]. Finally, other factors related to being infected by SARS-CoV-2, such as added anxiety, stress or isolation, might play a role in our findings. However, we did not find any non-imaging phenotype (measured pre-infection) that, when controlled for, greatly reduced the association of infection with the longitudinal IDP effects. It is also worth noting that most of the cases involved in this study were either asymptomatic or mild — and indeed, most of our significant olfactory-related results stand when excluding the infected participants with more moderate or severe COVID-

19 (hospitalised cases) — and most of the controls would have had also been exposed to higher levels of anxiety, stress and isolation during the pandemic in the UK. In addition, brain regions typically involved in these mental health factors, while overlapping with our limbic results, do not affect these regions consistently: higher and lower volumes have both been observed in these limbic structures for anxiety and stress, including higher (as opposed to lower in our study) volume of the parahippocampal gyrus in stress[68-71], while social isolation impacts a different network of brain regions (the so-called "default mode" network), which does not overlap with the pattern associated with SARS-CoV-2 infection in our results[72]. Whether our findings, which seem to delineate a clear limbic network of the primary and secondary olfactory and possibly gustatory cortex, reflect a corollary effect of the infection such as increased anxiety or isolation, an indirect brain-related pathophysiological process of SARS-CoV-2 through either anterograde degeneration or neuroinflammatory process, or a direct effect of the spread of the virus itself, remains to be elucidated.

Many of our results were found using imaging biomarkers of grey matter thickness or volume, which can be sensitive markers of a neurodegenerative process compared with other imaging modalities[73], and are robust measurements that makes them ideal in a longitudinal setting[74]. In fact, the longitudinal differences between the SARS-CoV-2 positive and control groups, while significantly localised in a limbic olfactory and gustatory network, seemed also — at a lower level — to be generalised, as illustrated in the significant shift in the distribution of Z values over the entire cortical surface (**Figure S7**). This means that there is an overall stronger decrease of grey matter thickness across the entire cortex in the infected participants, but that this effect is particularly dominant in the olfactory system. A marked atrophy of fronto-parietal and temporal regions can also be seen when contrasting hospitalised and non-hospitalised cases, suggesting increased damage in the more moderate and severe cases, with an additional significant shift in Z values (**Figure S7**). The pattern of the loss of grey matter in the hospitalised patients compared with the milder cases is consistent with PET FDG reports showing a fronto-parietal and temporal decrease in glucose in hospitalised COVID-19 patients [16,75].

The overlapping olfactory- and memory-related functions of the regions shown to alter significantly over time in SARS-CoV-2, including the parahippocampal gyrus/perirhinal cortex, entorhinal cortex and hippocampus in particular (**Tables S1 and S2**), raise the possibility that

longer-term consequences of SARS-CoV-2 infection might in time contribute to Alzheimer's disease or other forms of dementia[2]. This has led to the creation of an international consortium including the Alzheimer's Association and representatives from more than 30 countries to investigate these questions[2]. In particular, in our sample of infected participants with mainly mild symptoms, we found a significantly more pronounced increase in the duration taken to complete both trails of the Trail Making Test, which is known to be sensitive to detect, and discriminate, mild cognitive impairment and dementia from healthy ageing (**Figure 3**)[45]. In turn, the duration to complete the alphanumeric trail B correlated *post hoc* with the cognitive part of the cerebellum, namely crus II, which is also specifically activated by olfactory tasks[55,76]. It remains to be determined whether the loss of grey matter and increased tissue damage seen in memory-related regions of the brain may in turn increase the risk for these participants of developing dementia in the longer term[2,4,77].

Limitations of this study include the lack of stratification of severity of the cases, beyond the information of whether they had been hospitalised (information on O_2 saturation levels and details of treatment or hospital procedures is currently available on only a few participants), lack of clinical correlates (of particular relevance, potential hyposmic and hypogeusic symptoms and blood-based markers of inflammation), small number of participants from Asian, Black or other ethnic background other than White, some of the cases and controls' SARS-CoV-2 infection status identified through antibody lateral flow test kits that have varied diagnostic accuracy[78]. For those cases, no distinction is possible at present to determine whether a positive test is due to infection or vaccination, so potential cases identified only through lateral flow test in vaccinated participants were not included in this study. At present, information on the vaccination status (except for those identified through lateral flow test), and how both vaccination dates might interact with the date of infection, is unavailable. Another issue inherent to the recruitment strategy of UK Biobank, based on participants volunteering after being contacted at home for a possible re-imaging session, is the high number of mild cases. This can be seen however as a strength of this study: the majority of the brain imaging publications so far have focussed on moderate to severe cases of COVID-19[9], hence there is a fundamental need for more information on the cerebral effects of the disease in its milder form. The UK Biobank COVID-19 re-imaging study is ongoing, and further information will eventually be made available. Finally, on the imaging side, the UK Biobank

scanning protocol does not have the spatial resolution that would allow us to look at brainstem nuclei (with the exception of the substantia nigra), including those that are key autonomic and respiratory control centres.

This is the first longitudinal imaging study comparing brain scans acquired from individuals before and after SARS-CoV-2 infection to those scans from a well-matched control group. It also is one of the largest COVID-19 brain imaging studies, with 785 participants including 401 individuals infected by SARS-CoV-2. Its unique design makes it possible to more confidently tease apart the pathogenic contribution associated with the infection from pre-existing risk factors. By using automated, objective and quantitative methods, we uncovered a consistent spatial pattern of longitudinal abnormalities in limbic brain regions forming a mainly olfactory network. Whether these abnormal changes are the hallmark of the spread of the pathogenic effects, or of the virus itself in the brain, and whether these may prefigure a future vulnerability of the limbic system in particular, including memory, for these participants, remains to be investigated.

Acknowledgements

This work was primarily supported by a Wellcome Trust Collaborative Award 215573/Z/19/Z. KLM is supported by a Wellcome Trust Senior Research Fellowship 202788/Z/16/Z. The Wellcome Centre for Integrative Neuroimaging (WIN FMRIB) is supported by centre funding from the Wellcome Trust (203139/Z/16/Z). This research has been conducted in part using the UK Biobank Resource under Application Number 8107. We are grateful to UK Biobank for making the data available, and to all UK Biobank study participants, who generously donated their time to make this resource possible. We would like to thank Profs. Bruce Fischl and Doug Greve on guidance with the FreeSurfer analyses. Analysis was carried out at the Oxford Biomedical Research Computing (BMRC) facility. BMRC is a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute, supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre.

Ethics

Human subjects: UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) to obtain and disseminate data and samples from the participants (http://www. ukbiobank.ac.uk/ethics/), and these ethical regulations cover the work in this study. Written informed consent was obtained from all participants.

References

1 Paterson, R. W. *et al.* The emerging spectrum of COVID-19 neurology: clinical, radiological and laboratory findings. *Brain* **143**, 3104-3120, doi:10.1093/brain/awaa240 (2020).

2 de Erausquin, G. A. *et al.* The chronic neuropsychiatric sequelae of COVID-19: The need for a prospective study of viral impact on brain functioning. *Alzheimers Dement*, doi:10.1002/alz.12255 (2021).

3 Yang, A. C. *et al.* Dysregulation of brain and choroid plexus cell types in severe COVID-19. *Nature*, doi:10.1038/s41586-021-03710-0 (2021).

4 Deleidi, M. & Isacson, O. Viral and inflammatory triggers of neurodegenerative diseases. *Sci Transl Med* **4**, 121ps123, doi:10.1126/scitranslmed.3003492 (2012).

5 Butowt, R., Meunier, N., Bryche, B. & von Bartheld, C. S. The olfactory nerve is not a likely route to brain infection in COVID-19: a critical review of data from humans and animal models. *Acta Neuropathol* **141**, 809-822, doi:10.1007/s00401-021-02314-2 (2021).

6 Taquet, M., Geddes, J. R., Husain, M., Luciano, S. & Harrison, P. J. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry* **8**, 416-427, doi:10.1016/S2215-0366(21)00084-5 (2021).

7 Taquet, M., Luciano, S., Geddes, J. R. & Harrison, P. J. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry* **8**, 130-140, doi:10.1016/S2215-0366(20)30462-4 (2021).

8 Helms, J. *et al.* Neurologic Features in Severe SARS-CoV-2 Infection. *N Engl J Med* **382**, 2268-2270, doi:10.1056/NEJMc2008597 (2020).

9 Manca, R., De Marco, M., Ince, P. G. & Venneri, A. Heterogeneity in Regional Damage Detected by Neuroimaging and Neuropathological Studies in Older Adults With COVID-19: A Cognitive-Neuroscience Systematic Review to Inform the Long-Term Impact of the Virus on Neurocognitive Trajectories. *Front Aging Neurosci* **13**, 646908, doi:10.3389/fnagi.2021.646908 (2021).

10 Mukerji, S. S. & Solomon, I. H. What can we learn from brain autopsies in COVID-19? *Neuroscience letters* **742**, 135528, doi:10.1016/j.neulet.2020.135528 (2021).

11 Meinhardt, J. *et al.* Olfactory transmucosal SARS-CoV-2 invasion as a port of central nervous system entry in individuals with COVID-19. *Nat Neurosci* **24**, 168-175, doi:10.1038/s41593-020-00758-5 (2021).

12 Puelles, V. G. *et al.* Multiorgan and Renal Tropism of SARS-CoV-2. *N Engl J Med* **383**, 590-592, doi:10.1056/NEJMc2011400 (2020).

13 Matschke, J. *et al.* Neuropathology of patients with COVID-19 in Germany: a post-mortem case series. *Lancet Neurol* **19**, 919-929, doi:10.1016/S1474-4422(20)30308-2 (2020).

14 Lechien, J. R. *et al.* Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur Arch Otorhinolaryngol* **277**, 2251-2261, doi:10.1007/s00405-020-05965-1 (2020).

15 Cooper, K. W. *et al.* COVID-19 and the Chemical Senses: Supporting Players Take Center Stage. *Neuron* **107**, 219-233, doi:10.1016/j.neuron.2020.06.032 (2020).

16 Hosp, J. A. *et al.* Cognitive impairment and altered cerebral glucose metabolism in the subacute stage of COVID-19. *Brain* **144**, 1263-1276, doi:10.1093/brain/awab009 (2021).

17 Postma, E. M., Smeets, P. A. M., Boek, W. M. & Boesveldt, S. Investigating morphological changes in the brain in relation to etiology and duration of olfactory dysfunction with voxel-based morphometry. *Sci Rep* **11**, 12704, doi:10.1038/s41598-021-92224-w (2021).

18 Butowt, R. & Bilinska, K. SARS-CoV-2: Olfaction, Brain Infection, and the Urgent Need for Clinical Samples Allowing Earlier Virus Detection. *ACS Chem Neurosci* **11**, 1200-1203, doi:10.1021/acschemneuro.0c00172 (2020).

19 Netland, J., Meyerholz, D. K., Moore, S., Cassell, M. & Perlman, S. Severe acute respiratory syndrome coronavirus infection causes neuronal death in the absence of encephalitis in mice transgenic for human ACE2. *J Virol* **82**, 7264-7275, doi:10.1128/JVI.00737-08 (2008).

20 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**, doi:10.1126/sciadv.abc5801 (2020).

21 Carmichael, S. T., Clugnet, M. C. & Price, J. L. Central olfactory connections in the macaque monkey. *The Journal of comparative neurology* **346**, 403-434, doi:10.1002/cne.903460306 (1994).

22 Palouzier-Paulignan, B. *et al.* Olfaction under metabolic influences. *Chem Senses* **37**, 769-797, doi:10.1093/chemse/bjs059 (2012).

23 Guedj, E. *et al.* (18)F-FDG brain PET hypometabolism in post-SARS-CoV-2 infection: substrate for persistent/delayed disorders? *Eur J Nucl Med Mol Imaging* **48**, 592-595, doi:10.1007/s00259-020-04973-x (2021).

24 Raman, B. *et al.* Medium-term effects of SARS-CoV-2 infection on multiple vital organs, exercise capacity, cognition, quality of life and mental health, post-hospital discharge. *EClinicalMedicine* **31**, 100683, doi:10.1016/j.eclinm.2020.100683 (2021).

25 Reichert, J. L. & Schopf, V. Olfactory Loss and Regain: Lessons for Neuroplasticity. *Neuroscientist* **24**, 22-35, doi:10.1177/1073858417703910 (2018).

26 Han, P. *et al.* Olfactory brain gray matter volume reduction in patients with chronic rhinosinusitis. *Int Forum Allergy Rhinol* **7**, 551-556, doi:10.1002/alr.21922 (2017).

27 Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience* **19**, 1523-1536, doi:10.1038/nn.4393 (2016).

28 Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400-424, doi:10.1016/j.neuroimage.2017.10.034 (2018).

29 Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun* **11**, 2624, doi:10.1038/s41467-020-15948-9 (2020).

30 Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216, doi:10.1038/s41586-018-0571-7 (2018).

31 Pauli, W. M., Nili, A. N. & Tyszka, J. M. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci Data* **5**, 180063, doi:10.1038/sdata.2018.63 (2018).

32 Griffanti, L. *et al.* BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. *NeuroImage* **141**, 191-205, doi:10.1016/j.neuroimage.2016.07.018 (2016).

33 Wang, C. *et al.* Methods for quantitative susceptibility and R2* mapping in whole post-mortem brains at 7T applied to amyotrophic lateral sclerosis. *NeuroImage* **222**, 117216, doi:10.1016/j.neuroimage.2020.117216 (2020).

34 Iglesias, J. E. *et al.* A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *NeuroImage* **183**, 314-326, doi:10.1016/j.neuroimage.2018.08.012 (2018).

35 Iglesias, J. E. *et al.* Bayesian longitudinal segmentation of hippocampal substructures in brain MRI using subject-specific atlases. *NeuroImage* **141**, 542-555, doi:10.1016/j.neuroimage.2016.07.020 (2016).

36 Iglesias, J. E. *et al.* Bayesian segmentation of brainstem structures in MRI. *NeuroImage* **113**, 184-195, doi:10.1016/j.neuroimage.2015.02.065 (2015).

37 Saygin, Z. M. *et al.* High-resolution magnetic resonance imaging reveals nuclei of the human amygdala: manual segmentation to automatic atlas. *NeuroImage* **155**, 370-382, doi:10.1016/j.neuroimage.2017.04.046 (2017).

38 Avery, J. A. *et al.* Taste Quality Representation in the Human Brain. *J Neurosci* **40**, 1042-1052, doi:10.1523/JNEUROSCI.1751-19.2019 (2020).

39 Zhou, G., Lane, G., Cooper, S. L., Kahnt, T. & Zelano, C. Characterizing functional pathways of the human olfactory system. *Elife* **8**, doi:10.7554/eLife.47177 (2019).

40 Neudorfer, C. *et al.* A high-resolution in vivo magnetic resonance imaging atlas of the human hypothalamic region. *Sci Data* **7**, 305, doi:10.1038/s41597-020-00644-6 (2020).

41 Alfaro-Almagro, F. *et al.* Confound modelling in UK Biobank brain imaging. *NeuroImage* **224**, 117002, doi:10.1016/j.neuroimage.2020.117002 (2021).

42 Vickers, A. J. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol* **1**, 6, doi:10.1186/1471-2288-1-6 (2001).

43 Papst, I. *et al.* Age-dependence of healthcare interventions for COVID-19 in Ontario, Canada. *BMC Public Health* **21**, 706, doi:10.1186/s12889-021-10611-4 (2021).

44 Levin, A. T. *et al.* Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol* **35**, 1123-1138, doi:10.1007/s10654-020-00698-1 (2020).

45 Ashendorf, L. *et al.* Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Arch Clin Neuropsychol* **23**, 129-137, doi:10.1016/j.acn.2007.11.005 (2008).

46 Ryan, J. J. *et al.* The WASI matrix reasoning subtest: performance in traumatic brain injury, stroke, and dementia. *Int J Neurosci* **115**, 129-136, doi:10.1080/00207450490512704 (2005).

47 Tabrizi, S. J. *et al.* Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *The Lancet. Neurology* **11**, 42-53, doi:10.1016/S1474-4422(11)70263-0 (2012).

48 Salat, D. H. *et al.* Thinning of the cerebral cortex in aging. *Cerebral cortex* **14**, 721-730, doi:10.1093/cercor/bhh032 (2004).

49 Ronnlund, M., Nyberg, L., Backman, L. & Nilsson, L. G. Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychol Aging* **20**, 3-18, doi:10.1037/0882-7974.20.1.3 (2005).

50 Diana, R. A., Yonelinas, A. P. & Ranganath, C. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in cognitive sciences* **11**, 379-386, doi:10.1016/j.tics.2007.08.001 (2007).

51 Staresina, B. P., Duncan, K. D. & Davachi, L. Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *J Neurosci* **31**, 8739-8747, doi:10.1523/JNEUROSCI.4978-10.2011 (2011).

52 Naya, Y. & Suzuki, W. A. Integrating what and when across the primate medial temporal lobe. *Science (New York, N.Y* **333**, 773-776, doi:10.1126/science.1206773 (2011).

53 Doty, R. L. Olfaction: Smell of Change in the Air. *Cerebrum* **2017** (2017).

54 Chikazoe, J., Lee, D. H., Kriegeskorte, N. & Anderson, A. K. Distinct representations of basic taste qualities in human gustatory cortex. *Nat Commun* **10**, 1048, doi:10.1038/s41467-019-08857-z (2019).

55 Ferdon, S. & Murphy, C. The cerebellum and olfaction in the aging brain: a functional magnetic resonance imaging study. *NeuroImage* **20**, 12-21, doi:10.1016/s1053-8119(03)00276-3 (2003).

56 Kas, A. *et al.* The cerebral network of COVID-19-related encephalopathy: a longitudinal voxel-based 18F-FDG-PET study. *Eur J Nucl Med Mol Imaging* **48**, 2543-2557, doi:10.1007/s00259-020-05178-y (2021).

57 Qin, Y. *et al.* Long-term microstructure and cerebral blood flow changes in patients recovered from COVID-19 without neurological manifestations. *J Clin Invest* **131**, doi:10.1172/JCI147329 (2021).

58 Tsai, S. T., Lu, M. K., San, S. & Tsai, C. H. The Neurologic Manifestations of Coronavirus Disease 2019 Pandemic: A Systemic Review. *Front Neurol* **11**, 498, doi:10.3389/fneur.2020.00498 (2020).

59 Han, P., Musch, M., Abolmaali, N. & Hummel, T. Improved Odor Identification Ability and Increased Regional Gray Matter Volume After Olfactory Training in Patients With Idiopathic Olfactory Loss. *Iperception* **12**, 20416695211005811, doi:10.1177/20416695211005811 (2021).

60 Blazhenets, G. *et al.* Slow but evident recovery from neocortical dysfunction and cognitive impairment in a series of chronic COVID-19 patients. *J Nucl Med*, doi:10.2967/jnumed.121.262128 (2021).

61 Camara, M. L. *et al.* Effects of centrally administered etanercept on behavior, microglia, and astrocytes in mice following a peripheral immune challenge. *Neuropsychopharmacology* **40**, 502-512, doi:10.1038/npp.2014.199 (2015).

62 Klein, R. S., Garber, C. & Howard, N. Infectious immunity in the central nervous system and brain function. *Nat Immunol* **18**, 132-141, doi:10.1038/ni.3656 (2017).

63 Elmore, M. R. *et al.* Respiratory viral infection in neonatal piglets causes marked microglia activation in the hippocampus and deficits in spatial learning. *J Neurosci* **34**, 2120-2129, doi:10.1523/JNEUROSCI.2180-13.2014 (2014).

64 Ludlow, M. *et al.* Neurotropic virus infections as the cause of immediate and delayed neuropathology. *Acta Neuropathol* **131**, 159-184, doi:10.1007/s00401-015-1511-3 (2016).

65 Hosseini, S. *et al.* Long-Term Neuroinflammation Induced by Influenza A Virus Infection and the Impact on Hippocampal Neuron Morphology and Function. *J Neurosci* **38**, 3060-3080, doi:10.1523/JNEUROSCI.1740-17.2018 (2018).

66 Falcao, A. M. *et al.* The path from the choroid plexus to the subventricular zone: go with the flow! *Front Cell Neurosci* **6**, 34, doi:10.3389/fncel.2012.00034 (2012).

67 Chintamen, S., Imessadouene, F. & Kernie, S. G. Immune Regulation of Adult Neurogenic Niches in Health and Disease. *Front Cell Neurosci* **14**, 571071, doi:10.3389/fncel.2020.571071 (2020).

68 Wang, X., Cheng, B., Luo, Q., Qiu, L. & Wang, S. Gray Matter Structural Alterations in Social Anxiety Disorder: A Voxel-Based Meta-Analysis. *Front Psychiatry* **9**, 449, doi:10.3389/fpsyt.2018.00449 (2018).

69 Wang, X. *et al.* Distinct grey matter volume alterations in adult patients with panic disorder and social anxiety disorder: A systematic review and voxel-based morphometry meta-analysis. *J Affect Disord* **281**, 805-823, doi:10.1016/j.jad.2020.11.057 (2021).

70 Bromis, K., Calem, M., Reinders, A., Williams, S. C. R. & Kempton, M. J. Meta-Analysis of 89 Structural MRI Studies in Posttraumatic Stress Disorder and Comparison With Major Depressive Disorder. *Am J Psychiatry* **175**, 989-998, doi:10.1176/appi.ajp.2018.17111199 (2018).

71 Kuhn, S. & Gallinat, J. Gray matter correlates of posttraumatic stress disorder: a quantitative meta-analysis. *Biological psychiatry* **73**, 70-74, doi:10.1016/j.biopsych.2012.06.029 (2013).

72 Spreng, R. N. *et al.* The default network of the human brain is associated with perceived social isolation. *Nat Commun* **11**, 6393, doi:10.1038/s41467-020-20039-w (2020).

73 Douaud, G. *et al.* DTI measures in crossing-fibre areas: increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. *NeuroImage* **55**, 880-890, doi:10.1016/j.neuroimage.2010.12.008 (2011).

74 Douaud, G. *et al.* Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9523-9528, doi:10.1073/pnas.1301816110 (2013).

75 Blazhenets, G. *et al.* Slow but Evident Recovery from Neocortical Dysfunction and Cognitive Impairment in a Series of Chronic COVID-19 Patients. *J Nucl Med* **62**, 910-915, doi:10.2967/jnumed.121.262128 (2021).

76 Sobel, N. *et al.* Odorant-induced and sniff-induced activation in the cerebellum of the human. *J Neurosci* **18**, 8990-9001 (1998).

77 Heneka, M. T., Kummer, M. P. & Latz, E. Innate immune activation in neurodegenerative disease. *Nat Rev Immunol* **14**, 463-477, doi:10.1038/nri3705 (2014).

78 Deeks, J. J. *et al.* Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev* **6**, CD013652, doi:10.1002/14651858.CD013652 (2020).

Supplementary Material

Figure S1. Age distributions for SARS-CoV-2 positive participants and controls at each timepoint do not differ significantly. Two-sample Kolmogorov-Smirnov was used to compute the P values for age comparisons, since age for each group was not normally distributed (Lilliefors P = 1e-03 for each group, and both age at Scan 1 or Scan 2). This showed no significant difference in age distribution between SARS-CoV-2 participants and controls at Scan 1: P = 0.15 or at Scan 2: P = 0.08.

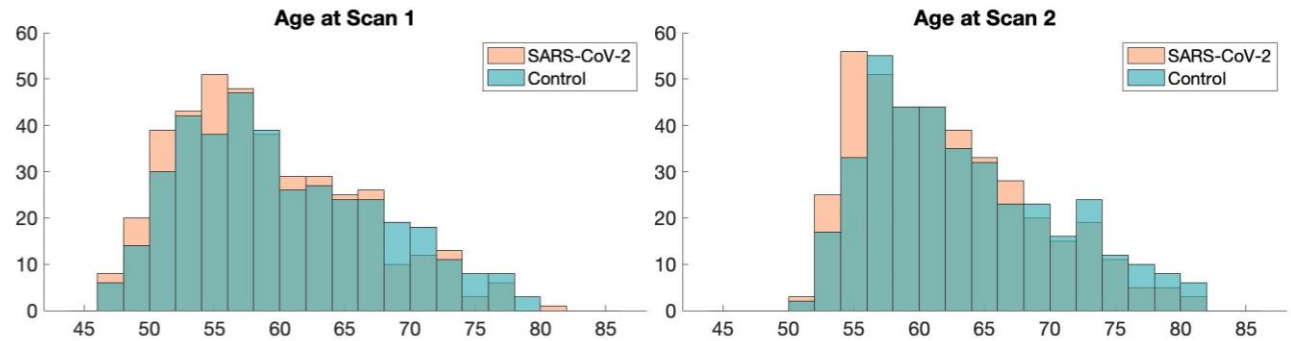

Figure S2. Histograms showing the well-matched distributions of Scan 1 - Scan 2 intervals for case and control groups. The below IDP reproducibility **Figure S3** shows, for comparison against the cases and controls, reproducibility from around 3,000 (2,943) UK Biobank participants who had returned for a second scan prior to the pandemic; hence we also show here the interscan intervals for this "3k" group, with tighter control over this interval.

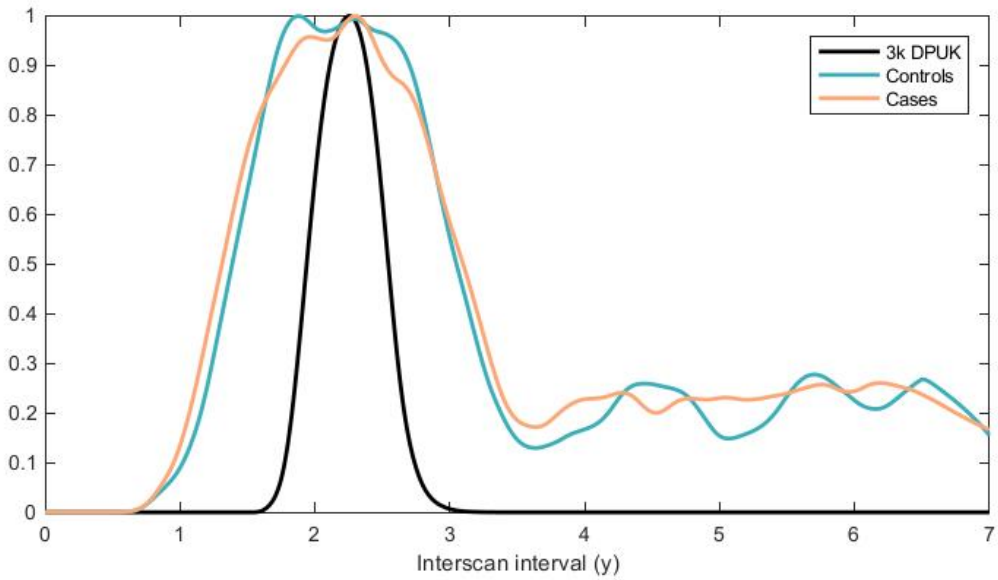

Figure S3. Scan-rescan reproducibility for all 2,047 IDPs used in the main modelling. Each dot represents a single IDP, arranged into different classes of IDPs. For each IDP, the vector of values for each subject (i.e., 785x1 vector) from the first scan was correlated with the equivalent vector of IDP values from the second scan. The y axis shows the resulting correlation coefficient. These calculations are made separately for the pre-pandemic scan-rescan datasets ("3k DPUK"), and for cases and controls, demonstrating highly similar distributions within each IDP class for all 3 subject groups.

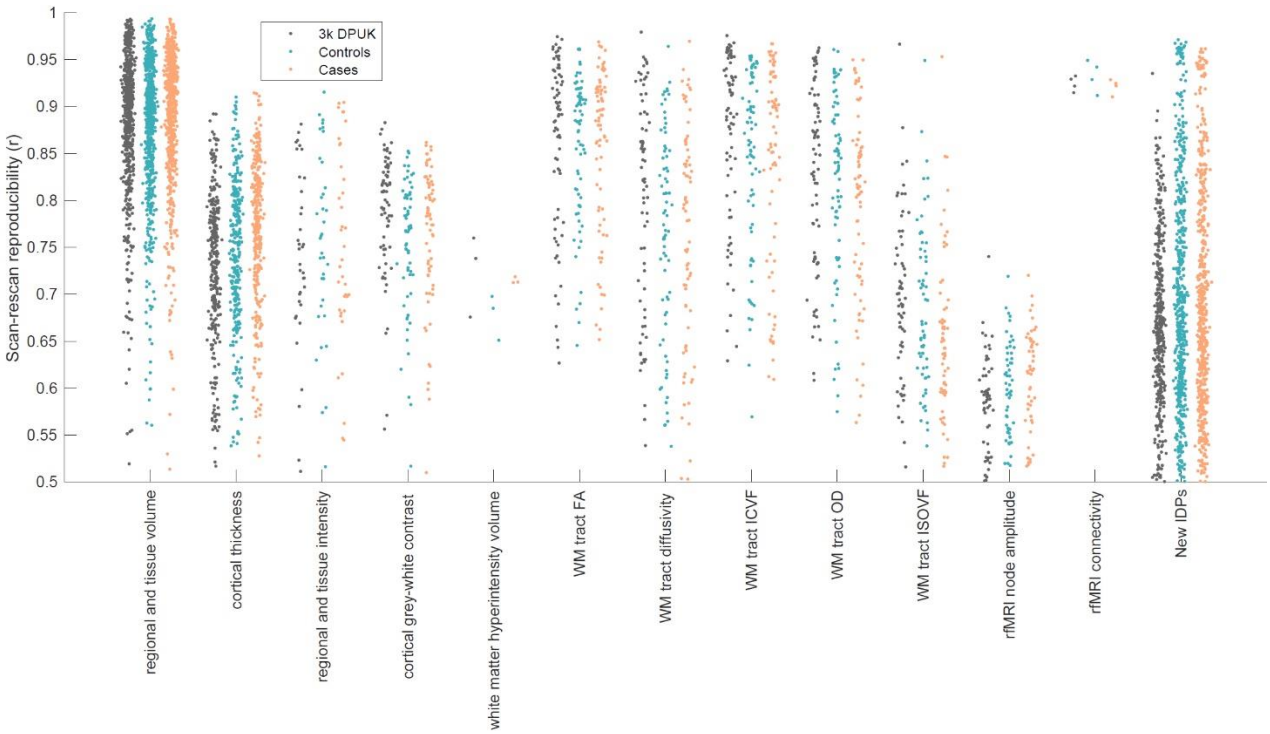

Figure S4. QQ plot for -Log$_{10}$(P$_{uncorrected}$) against the theoretical null distribution. The black line at y=x shows the expected plot if no effects were present in the data. Orange points reflect ΔIDPs where the case-control effect passes FDR significance, and blue reflects those that do not.

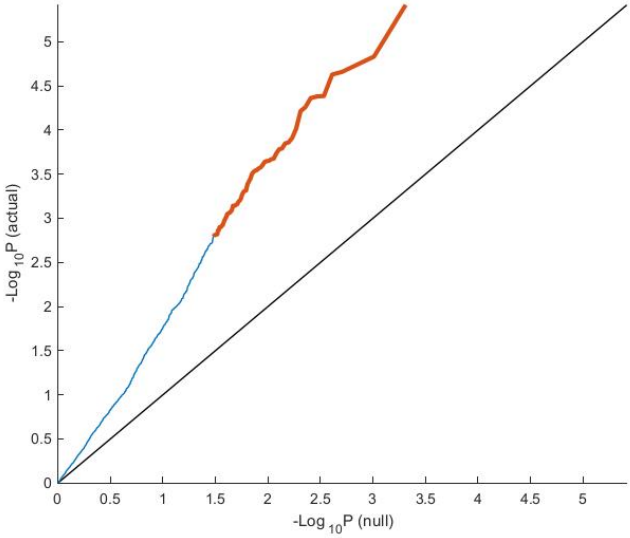

Figure S5. Model Z-statistics (one point per IDP, arranged in IDP classes) for the 4 main models. Note that these are model Z-statistics, not raw effect size. Some IDP classes (e.g., cortical thickness and grey-white intensity contrast) show consistent group-difference effect directions across most IDPs (i.e., different brain regions), and all 4 models.

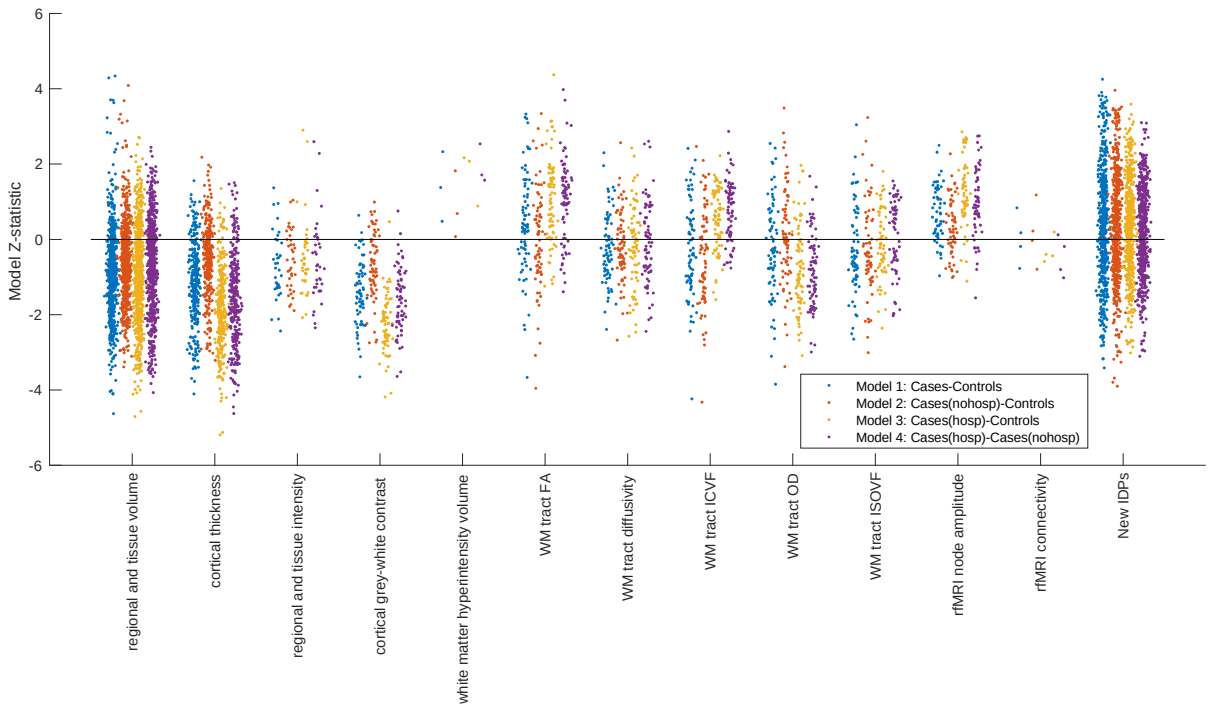

Figure S6. Examples of some of the most significant longitudinal group comparison results - exploratory approach. Four amongst the top IDPs consistently showing longitudinal differences between SARS-CoV-2 cases and controls. All demonstrate either a greater reduction in local or global brain thickness and volume, or an increase in CSF volume. For each four IDP are the percentage changes with age for the two groups, obtained by normalising ΔIDP using as baseline the values for the corresponding IDPs across up to 46,743 scans from the larger UK Biobank imaging study (created using a 10 year sliding window across cases and controls, with standard errors in grey). The somewhat counterintuitive increase in thickness in the rostral anterior cingulate cortex in older controls has been previously consistently reported in studies of ageing, together with that of the orbitofrontal cortex[48].

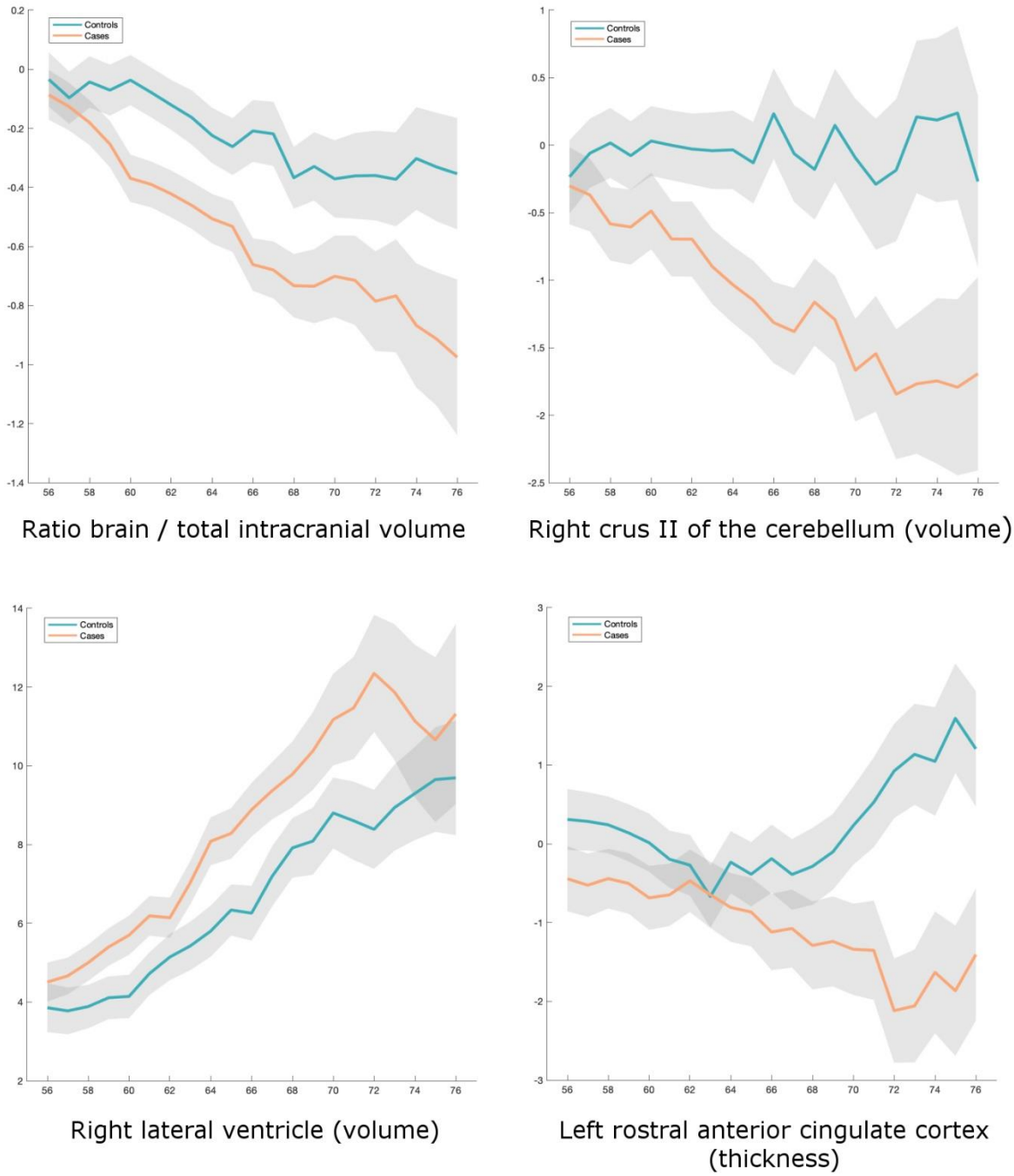

Figure S7. Histograms of group comparison Z statistics of longitudinal change in cortical thickness. Top row. Model 1: All SARS-CoV-2 participants vs controls. Left, histogram of Z-statistics (blue) across all cortical vertices of the left hemisphere (with grey lines showing 100 null histograms created through random permutations of the group variable). Middle: right hemisphere Z-statistics (orange) and matched nulls. Right: the same left and right hemisphere Z-statistics histograms overlaid (with a pooled null histogram in grey). **Bottom row.** Model 4: All hospitalised vs non-hospitalised SARS-CoV-2 participants. Same representation of Z-statistics histograms as in the top row.

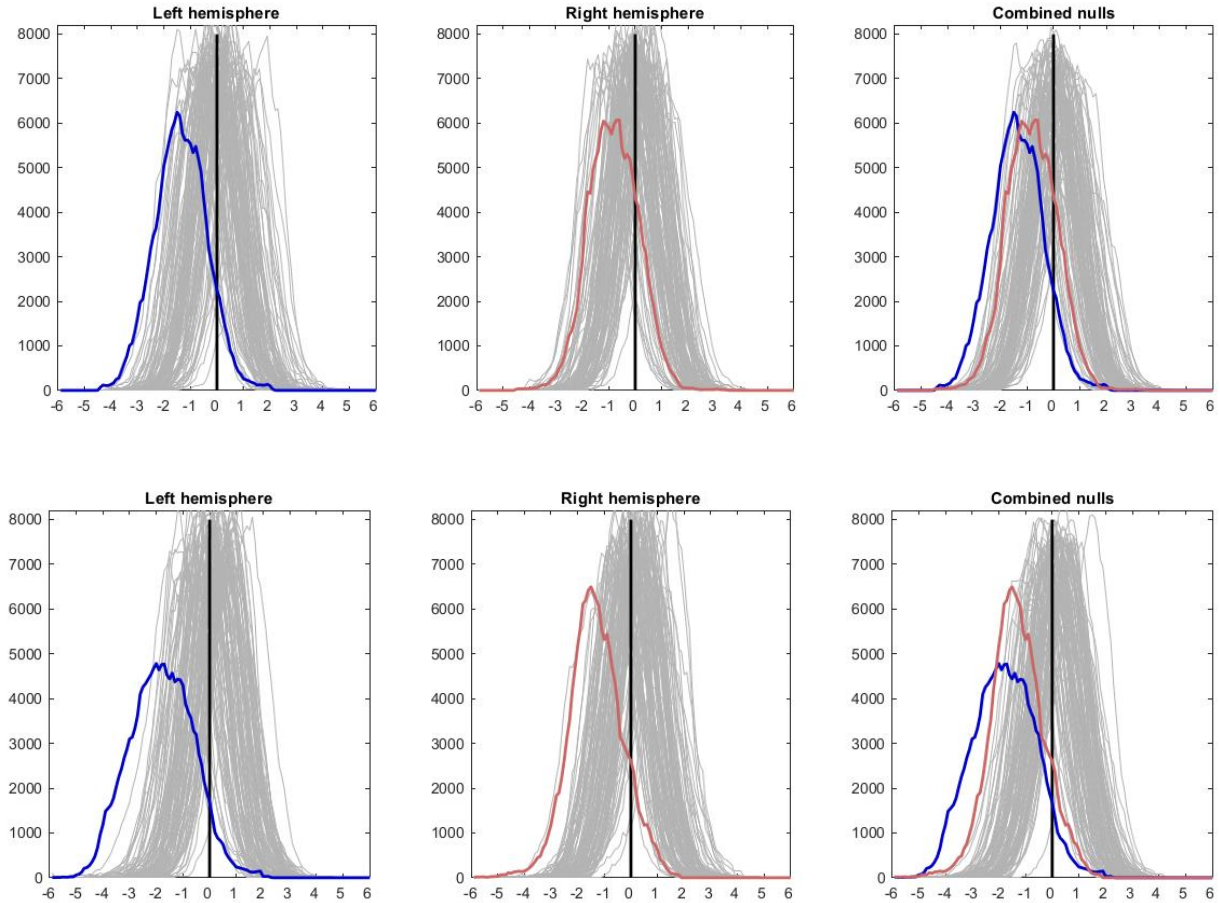

Table S1. Full list of reproducible IDPs used in the hypothesis-driven approach, and corresponding statistics for the longitudinal analyses (Models 1-4).

Please see additional attached PDF.

Table S2. Full list of reproducible IDPs used in the exploratory approach, and corresponding statistics for the longitudinal analyses (Models 1-4).

Please see additional attached PDF.

Table S3. Full list of reproducible IDPs used in the exploratory approach, and corresponding statistics for the cross-sectional, baseline analysis comparing SARS-CoV-2 and control groups.

Please see additional attached PDF.

Table S4. Full list of reproducible IDPs used in the exploratory approach, and corresponding statistics for the cross-sectional, second timepoint analysis comparing SARS-CoV-2 and control groups.

Please see additional attached PDF.