Data supplement to Su et al. Neuroinflammatory and morphological changes in late-life depression: the NIMROD study. Br J Psychiatry doi: 10.1192/bjp.bp.116.190165

Supplement DS1

METHOD

MRI guided PET Image Analysis

To estimate TSPO binding site density, non-displaceable binding potential (BP_{ND}) - a measure of specific binding - was determined using simplified reference tissue modelling of the dynamic [^{11}C]PK11195 PET data, with supervised cluster analysis to determine the reference tissue time-activity curve.[1] BP_{ND} was estimated voxel-wise and also for a set of regions based on the Hammers atlas (http://brain-development.org/brain-atlases/). The regions of interest were non-rigidly inverse-normalized to subjects' native T1 space, to which the dynamic PET image data were coregistered. Prior to the determination of BP_{ND}, regional PET data were corrected for cerebrospinal fluid (CSF) contamination using SPM tissue class segmentation to reduce CSF-induced partial volume effects, particularly those resulting from differential brain atrophy across subjects. To illustrate the spatial distribution of BP_{ND} in each individual depressed subject, [^{11}C]PK11195 BP_{ND} maps of both depressed and control groups were normalized to MNI space (http://www.bic.mni.mcgill.ca/ServicesAtlases/HomePage). Then, the BP_{ND} was transferred to Z-score with respect to the control group after non-brain areas (identified as not white or grey matter) were removed using SPM tissue class segments.

T2 FLAIR: Segmentation and Quantification of White Matter Hyperintensities

Based on T2 FLAIR scans, individual white matter hyperintensity (WMH) maps were segmented using the Lesion Segmentation Toolbox in Statistical Parametric Mapping 12 (SPM: Wellcome Institute of Neurology, University College London, UK). The technical details have been previously described[2] and this method has been shown to yield a high agreement with semi-manual lesion

volume estimation.[3] After the segmentation of lesion maps, a trained operator (S.G.) manually inspected and corrected the segmentations while blinded to group diagnosis. An in-house script[4] was then used to quantify periventricular (PvWMH), and deep WMH volumes, as well as total WMH. In addition, lobar WMH in bilateral frontal, temporal, parietal, and occipital regions were also calculated.

T1 Weighted: Voxel-based Morphometry and ROI Analysis of Grey Matter Volumes

Individual T1-weighted volumetric images were processed using an optimised voxel-based morphometry (VBM) (http://dbm.neuro.uni-jena.de) within SPM. T1-weighted images were first bias-corrected and segmented into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) partitions, and then affine normalised into MNI space. GM probability maps were modulated (intensity corrected for local volumetric changes induced during normalisation), followed by smoothing with an 8-mm full width at half maximum kernel. VBM analysis of GM differences was performed with a two-sample *T-test*, controlling for Total Intracranial Volume (TIV: calculated as the sum of total volumes of the GM, WM and CSF partitions). Analyses were restricted to the GM mask of the template, thresholded at 10% GM probability. Results were assessed at a statistical threshold of $p<0.001$ uncorrected. We did not perform a conservative multiple comparison correction given the small sample, particularly in the LLD group. From the modulated GM images, we used the Hammers atlas,[5] modified in-house to achieve consistency with [^{11}C]PK11195 PET parcellations, to extract the average volumes (mm^3) of specific regions of interests in the medial temporal lobe and the cingulate regions. Between-group differences in volumetric data were compared using ANCOVA while correcting for TIV.

T2 Weighted: Hippocampal Subfields Analysis

Hippocampal subfields were segmented manually using the high resolution (0.4 mm x 0.4 mm) T2 weighted MRI data. In order to increase the signal to noise ratio we acquired two images for each subject and aligned them before averaging to produce a final image. The averaged images for each

subject were then analysed by two raters (L.H. and S.S.), blinded to the diagnosis and subject characteristics, using an established manual approach, in which the hippocampi and subfields CA1, CA2, CA3/DG were traced on the 3 coronal slices directly posterior to the head of hippocampus;[6,7] see Fig. DS1.

The subiculum thickness (see Fig. DS1) was measured at the point immediately before it joined the medial hippocampus on the three slices that had hippocampal traces made. The entorhinal cortex (ERC) was measured at the same point on the first slice of the hippocampal tracings and then the two slices anterior to this (in the hippocampal head) with the line drawn perpendicular to the inferior border of the ERC. The values for the subfield area, the ERC and subiculum thickness were averaged across the left and right hemispheres and the three image slices analysed.

We assessed the intra-rater reliability of the tracings by repeating the measurements with the same rater one month after the initial tracings. It was performed on a test dataset of 3 AD, 3 MCI and 3 control subjects. (One control subject in the test dataset is a subject in our control group.) As with the initial tracing, the rater was blinded to the diagnosis when repeating the procedure a month later. For inter-rater reliability, both raters traced the same test dataset of 9 subjects. We measured the percentage difference in measurements and Intraclass Correlation Coefficient (ICC), using a two-way random model for absolute agreement of measurements.

Individual level Monte Carlo randomisation test for regional [^{11}C]PK11195 PET data

In clinical practice, it is essential to obtain information about each individual patient. So, in addition to the group comparison, we have performed an additional individual level analysis comparing each subject to the rest of the population. This analysis took advantage of the relatively large control group (N=13) and derived a set of *p*-values for every subject including both controls and LLD subjects. This kind of randomisation test was widely used in functional MRI analysis (such as those statistical

algorithms implemented in FSL[7]) and to the analysis of EEG data in a lie detection scenario, in which discriminating guilty information at single subject level is particularly critical[8,9]. The analysis procedure was as the follows:

1. Leave one subject *S* out;
2. Compute the mean BP_{ND} of the opposite group to *S*;
3. Compute the distance (or difference) between the mean BP_{ND} and the BP_{ND} of subject *S*; call this distance *d*;
4. Randomly permute the group labels among all subjects except for *S*. This is to simulate the null hypothesis in which the BP_{ND} for controls and patients would be sampled from the same distribution, i.e. group labels can be arbitrarily exchanged.
5. Compute the mean of the opposite group based on the permuted data, and the distance between the mean and the subject *S*. Call the new distance *ds*;
6. Repeat step 4 and step 5 for 10,000 times to build a null distribution of *ds*; (See Fig. DS2 for an example null distribution for parahippocampus ROI in subject LLD5.)
7. The percentage of times when *ds* is larger than *d* shows the likelihood of false positive, so can be converted to a *p*-value for subject *S*.

We have performed the above test for both subgenual anterior cingulate and parahippocampus ROIs, and the two *p*-values were further combined using the Stouffer's methods into a single *p*-value for each subject.

Supplement DS2

RESULTS

Subject Characteristics and Clinical Measurements

Depressed subjects (LLD) and controls (Con) did not differ in age, sex ratio, education or global cognition but had significantly higher blood CRP levels than controls, and a trend in Montgomery–Åsberg Depression Rating Scale (MADRS). (See Table DS1.) Table DS2 shows the previous history of depressive episodes for the LLD subjects.

White Matter Lesions

We found marginally increased WMH volumes in subjects with LLD in both periventricular and deep white matter. In terms of lobar distribution, the majority of WMH were in frontal and parietal regions, with temporal and occipital areas relatively spared compared with controls; see Table DS3. However, we found that white matter lesions were common in both controls and LLD subjects largely due to the age range of our subjects.

Grey Matter Volume

In the whole brain voxel wise analysis, we found bilateral hippocampal, right fusiform, right frontal and bilateral precuneus atrophy in LLD subjects compared with controls (controlled for age, sex, years of education and TIV but uncorrected at $p < 0.001$; not significant after FWE/FDR corrections); see Fig. DS3. In the ROI analysis using the modified Hammers atlas, we found reduction in GM volume in right hippocampal in LLD (controlled for age, sex, years of education and TIV but uncorrected for multiple comparisons); see Table DS4.

Hippocampal Segmentation Test–retest Reliability

The reliability of intra- and inter-rater measurements was comparable to what is found in the literature,[5,6] with differences in the CA1, CA2 and CA3/DG areas and subiculum and ERC thickness

below 13%, which is regarded as highly consistent. The intra-rater reliability was generally better than the inter-rater reliability in terms of both percentage size differences and ICC; see Table DS5.

Hippocampal Subfields

Using T2 weighted scans optimised for hippocampal regions, we found significant atrophy in CA1 and subiculum areas in the medial temporal lobe, and a marginally atrophic CA2 in subjects with LLD compared with controls, see Table DS6. This is consistent with both our GM volume analysis based on T1 weighted scans and prior research.

Individual level results for [^{11}C]PK11195 BP_{ND}

Using the Monte Carlo methods, we found that all LLD subjects have significantly increased [^{11}C]PK11195 BP_{ND} in subgenual anterior cingulate cortex and parahippocampus, while all controls did not have such an effect. Individual level *p*-values for five LLD subjects are shown in Table DS7. This result based on a more rigorous nonparametric statistical test that makes minimum assumption about the data showed a consistent finding with the group level inference reported in the main text. So, we can be more confident that the association in [^{11}C]PK11195 BP_{ND} was a real and robust effect, because the randomisation test when sufficient resamplings are taken, accurately approximates exhaustive permutation tests, which are, in a specific sense, statistically exact. The randomisation procedure has replicated this experiment a large number of times (10,000 in this case), providing what is effectively a large sample size when accumulating across replications and the effectiveness of the method at the individual-level carries over to a larger sample.

Age and Disease Duration Effects

Although not significant, LLD subjects were 5 years older than controls on average. Due to the small sample size, we could not remove the younger controls to match the age between groups, or control age effectively in the statistical test. However, we found no correlation between age and the

[^{11}C]PK11195 BP_{ND} in subgenual anterior cingulate cortex and parahippocampus for controls, suggesting the group difference was unlikely a result of age.

We found a trend level correlation between [^{11}C]PK11195 BP_{ND} in subgenual anterior cingulate cortex and age at disease onset (r=0.8, p=0.1) for subjects with LLD; see Fig. DS4. However, this relationship was not found in parahippocampus.

References

1. Yaqub M, van Berckel BN, Schuitemaker A, Hinz R, Turkheimer FE, Tomasi G, et al. Optimization of supervised cluster analysis for extracting reference tissue input curves in (R)-[(11)C]PK11195 brain PET studies. *J Cereb Blood Flow Metab* 2012; **32**: 1600–8.
2. Schmidt P, Gaser C, Arsic M, Buck D, Forschler A, Berthele A, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage* 2012; **59**:3774–3783.
3. Pareto D, Sastre-Garriga J, Aymerich FX, Auger C, Tintoré M, Montalban X, Rovira A. Lesion filling effect in regional brain volume estimations: a study in multiple sclerosis patients with low lesion load. *Neuroradiology*. 2016; **58**(5):467–474.
4. Firbank M.J., Minett T., O'Brien J.T. Changes in DWI and MRS associated with white matter hyperintensities in elderly subjects. *Neurology*. 2003; **61**:950–954.
5. Hammers A, Allom R, Koepp MJ, Free SL, Myers R, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* 2003; **19**(4):224–247.
6. Mueller SG, Stables L, Du AT, Schuff N, Truran D, Cashdollar N, Weiner MW. Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4T. *Neurobiol Aging*. 2007; **28**:719-726.

7. Firbank MJ, Blamire AM, Teodorczuk A, Teper E, Burton EJ, Mitra D, and O'Brien JT. High Resolution Imaging of the Medial Temporal Lobe in Alzheimer's disease and Dementia with Lewy Bodies. *Journal of Alzheimer's disease*. 2010; *21*:1129–1140.
8. Nichols T.E. and Holmes A.P. Nonparametric Permutation Tests for Functional Neuroimaging: A Primer with Examples. *Human Brain Mapping*, 2001; 15:1-25.
9. Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A. and Wyble, B. Subliminal Salient Search Illustrated: Deception Detected on the Fringe of Awareness, *PLoS One* 2013; 8(1): e54258, doi:10.1371/ journal.pone.0054258
10. Bowman, H., Filetti, M., Alsufyani, A., Janssen, D. and Su, L. Countering countermeasures: detecting identity lies by detecting conscious breakthrough, *PLoS One*, 2014; 9(3): e90595. doi:10.1371/journal.pone.0090595

	Controls	LLD	*p*
Age	68.0 (5.5)	73.2 (5.3)	0.108
Sex (F:M)	8:5	3:2	0.952
Education (yrs)	14.1 (2.8)	12.0 (3.9)	0.165
MMSE	28.7 (1.0)	27.6 (1.9)	0.275
ACE-R	91.3 (5.5)	88.4 (7.8)	0.547
CRP (mg/L)	1.2 (1.2)	8.0 (4.2)	**0.004**
ESR	12.1 (11.1)	8.3 (5.8)	0.592
WBC	6.9 (1.8)	7.7 (1.1)	0.178
GDS	1.5 (1.3)	3.0 (4.5)	0.896
HADS Anxiety	5.0 (4.0)	7.2 (4.2)	0.279
HADS Depression	2.4 (1.9)	3.0 (3.1)	0.747
MADRS	4.0 (4.6)	10.0 (7.1)	**0.065**

Table DS1. Demographic and clinical characteristic in mean (SD). P values were derived using nonparametric Mann-Whitney U test except for sex, which used the Chi-square test. Abbreviations: LLD - Late life depression; MMSE - Mini Mental State Examination; ACE-R – Addenbrooke's Cognitive Examination Revised; CRP - C-reactive protein; ESR – erythrocyte sedimentation rate; WBC – white blood cell count; GDS - Geriatric Depression Scale; HADS – Hospital Anxiety and Depression Scale; MADRS - Montgomery–Åsberg Depression Rating Scale.

	LLD1	LLD2	LLD3	LLD4	LLD5
Age of onset of first depression (years)	52	62	75	71	73
Years from disease onset	13	9	3	5	3
Number of previous depressive episodes	2	4	1	3	5

Table DS2. History of depressive episodes.

Volumes (mm^3)	Controls	LLD	*p*
Total WMH	4.47 (2.48)	9.18 (6.00)	0.12
Periventricular WMH	3.74 (2.06)	7.33 (4.45)	**0.059**
Deep WMH	0.73 (0.61)	1.85 (1.76)	**0.095**
Frontal L	0.95 (0.54)	1.92 (1.08)	**0.075**
Frontal R	1.03 (0.72)	2.29 (1.45)	**0.024**
Parietal L	0.32 (0.27)	1.07 (1.03)	**0.010**
Parietal R	0.39 (0.58)	0.95 (1.07)	**0.046**
Occipital L	0.30 (0.25)	0.29 (0.09)	0.94
Occipital R	0.28 (0.26)	0.31 (0.17)	0.70
Temporal L	0.27 (0.19)	0.51 (0.44)	0.25
Temporal R	0.23 (0.22)	0.50 (0.48)	0.21

Table DS3. WMH in periventricular and deep WM, as well as it's lobar distribution.

Abbreviations: LLD - Late life depression; L - Left; R - Right; WMH - white matter hyperintensity.

Volumes (mm^3)	Controls	LLD	p [a]	p [b]
Hippocampus L	1812.73 (136.73)	1739.55 (140.97)	0.170	0.646
Hippocampus R	2071.32 (132.76)	1959.99 (167.57)	**0.034***	0.288
Parahippocampal Gyrus L	2757.66 (253.34)	2840.81 (107.70)	0.586	0.607
Parahippocampal Gyrus R	2739.44 (252.60)	2602.34 (307.42)	0.101	0.685
Presubgenual Anterior Cingulate L	606.77 (108.99)	598.89 (54.33)	0.629	0.500
Presubgenual Anterior Cingulate R	499.24 (132.83)	492.29 (81.51)	0.743	0.773
Subgenual Anterior Cingulate L	700.09 (127.23)	669.03 (30.27)	0.362	0.305
Subgenual Anterior Cingulate R	612.97 (121.59)	603.88 (48.08)	0.657	0.444
Posterior Cingulate L	4203.35 (524.93)	4342.61(419.52)	0.744	0.864
Posterior Cingulate R	4091.32 (590.70)	4119.15 (363.35)	0.794	0.752
Insula L	6958.11 (798.16)	7316.50 (457.14)	0.412	0.405
Insula R	7124.87 (742.94)	7552.37 (728.02)	0.336	0.279
TIV	1501.8 (149.48)	1534.24 (165.60)	0.694	0.754

Table DS4. Volumetric comparisons of regions of interest between controls and late-life depression in mean (SD). p [a] Statistical comparisons were performed with ANCOVA accounting for TIV; p [b] ANCOVA accounting for TIV, age, gender, and education. Abbreviations: LLD - Late life depression; L - Left; R - Right; TIV - Total intracranial volume.

	CA1	CA2	CA3/DG	Subiculum thickness	ERC thickness
Inter-rater Reliability					
Percentage difference (%)	11.9	11.7	12.8	11.6	9.2
ICC (average measures)	0.775	0.798	0.629	0.751	0.572
Significance p value	< 0.001	< 0.001	0.018	< 0.001	0.01
Intra-rater Reliability					
Percentage difference (%)	8.3	6.5	9.3	8.3	4.0
ICC (average measures)	0.83	0.832	0.813	0.839	0.559
Significance p value	< 0.001	0.002	0.005	0.004	0.107

Table DS5. Hippocampal subfield segmentation reliability: Repeated measures for a total 9 samples (3 AD, 3 MCI and 3 Controls), percentage difference and ICC is a two-way random model testing absolute agreement. Abbreviations: ICC - Intraclass Correlation Coefficient.

	Controls	LLD	*p*
CA1 area (mm^3)	24.90 (2.92)	22.09 (0.41)	**0.019**
CA2 area (mm^3)	1.35 (0.23)	1.41 (0.36)	**0.08**
CA3/DG area (mm^3)	18.75 (3.25)	17.25 (1.32)	0.51
Subiculum thickness (mm)	1.95 (0.13)	1.73 (0.09)	**0.0039**
ERC thickness (mm)	2.34 (0.34)	2.24 (0.23)	0.8

Table DS6. Hippocampal subfield volumes and thicknesses - mean (SD). Abbreviations: LLD - Late life depression; CA - Cornu Ammonis.

Subject	*p*-value for parahippocampus	*p*-value for subgenual ACC	Combined *p*-value
LLD1	0.049	0.0042	**0.0012**
LLD2	0.0023	0.01	**0.00014**
LLD3	0.045	0.018	**0.0036**
LLD4	0.02	0.03	**0.0026**
LLD5	0.044	0.038	**0.0069**

Table DS7. Individual level *p*-values for five LLD subjects based on Monte Carlo Randomisation test. Both the ROI and combined p-values are significant for all LLD subjects revealing an overall 100% hit rate.

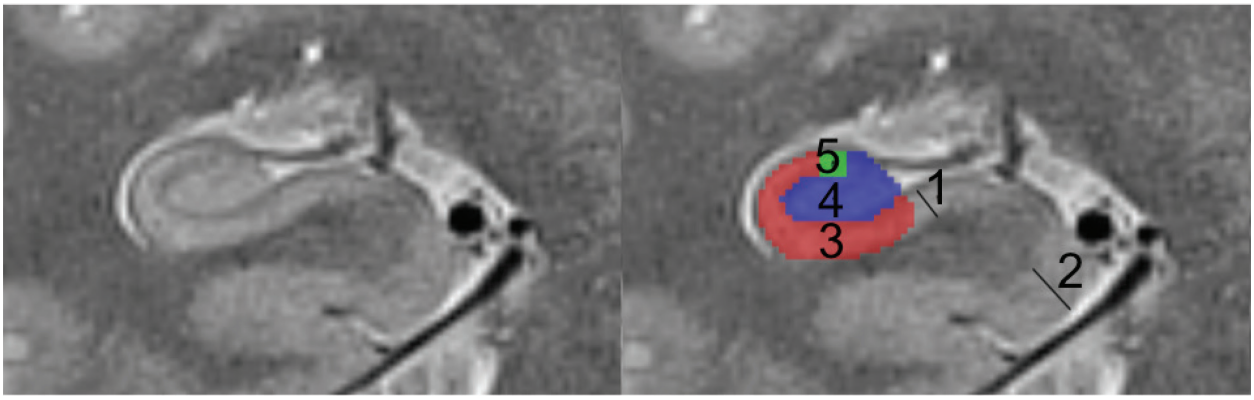

Fig DS1 Manual segmentation of the hippocampal subfields: 1. Subiculum, delineated at the medial border of the hippocampus, 2. Entorhinal cortex (ERC), measured perpendicular to the inferior border, 3. Cornu Ammonis 1 (CA1) subfield, 4. Cornu Ammonis 3/dentate gyrus (CA3/DG) subfield: the CA1-CA3 boundary is delineated using the clearly visible hypointense line (left image) 5. Cornu Ammonis 2 (CA2) subfield, marked as the height of the CA1 subfield at approximately the midpoint of horizontal axis.

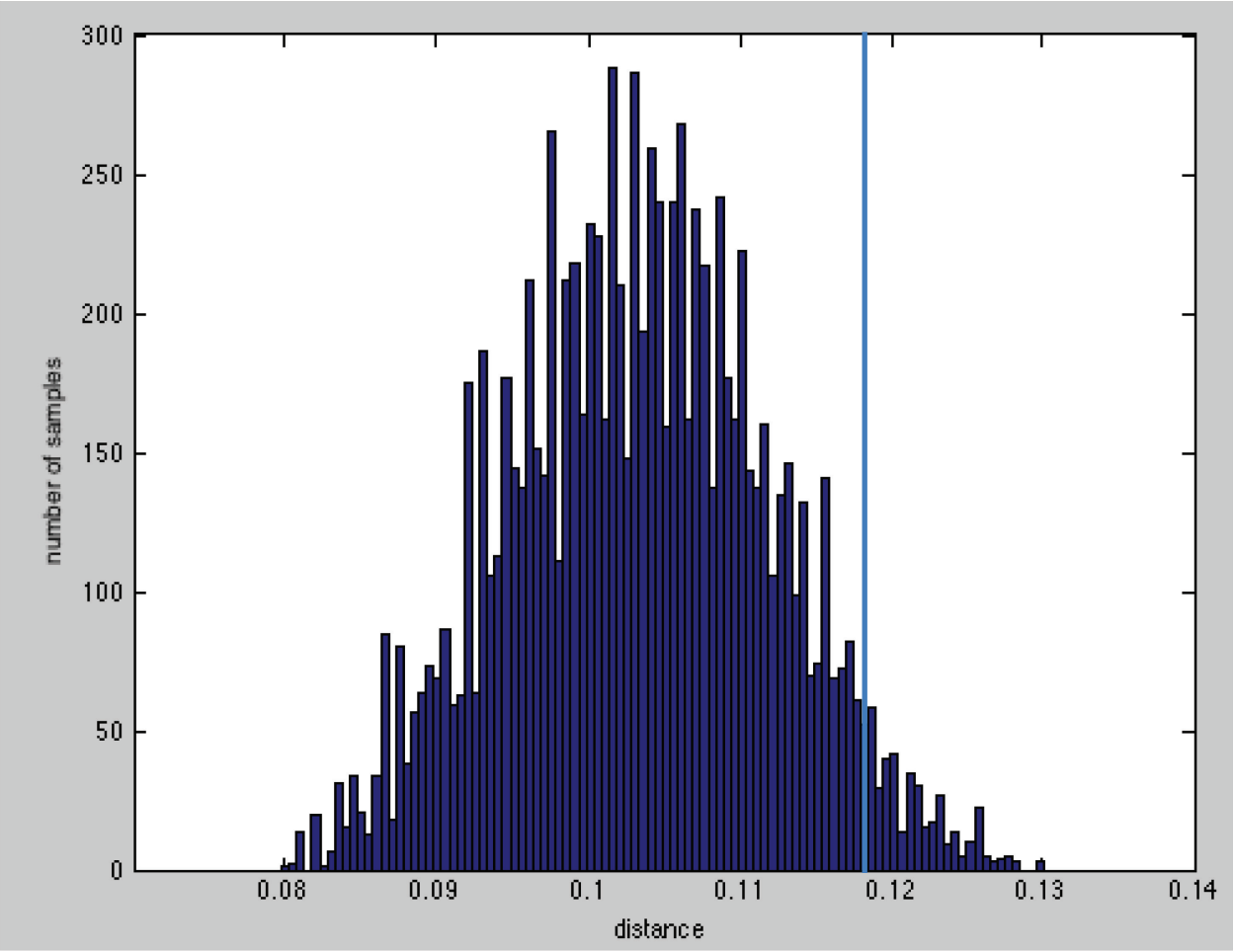

Fig. DS2 Distance between individual sample and population mean null hypothesis distributions. Randomisation inferred null hypothesis distributions for differences between a LLD subject and the sample mean of 13 surrogate control subjects generated from permutation. The vertical line marks true observed value for the LLD subject and *p*-value region to the right of the line.

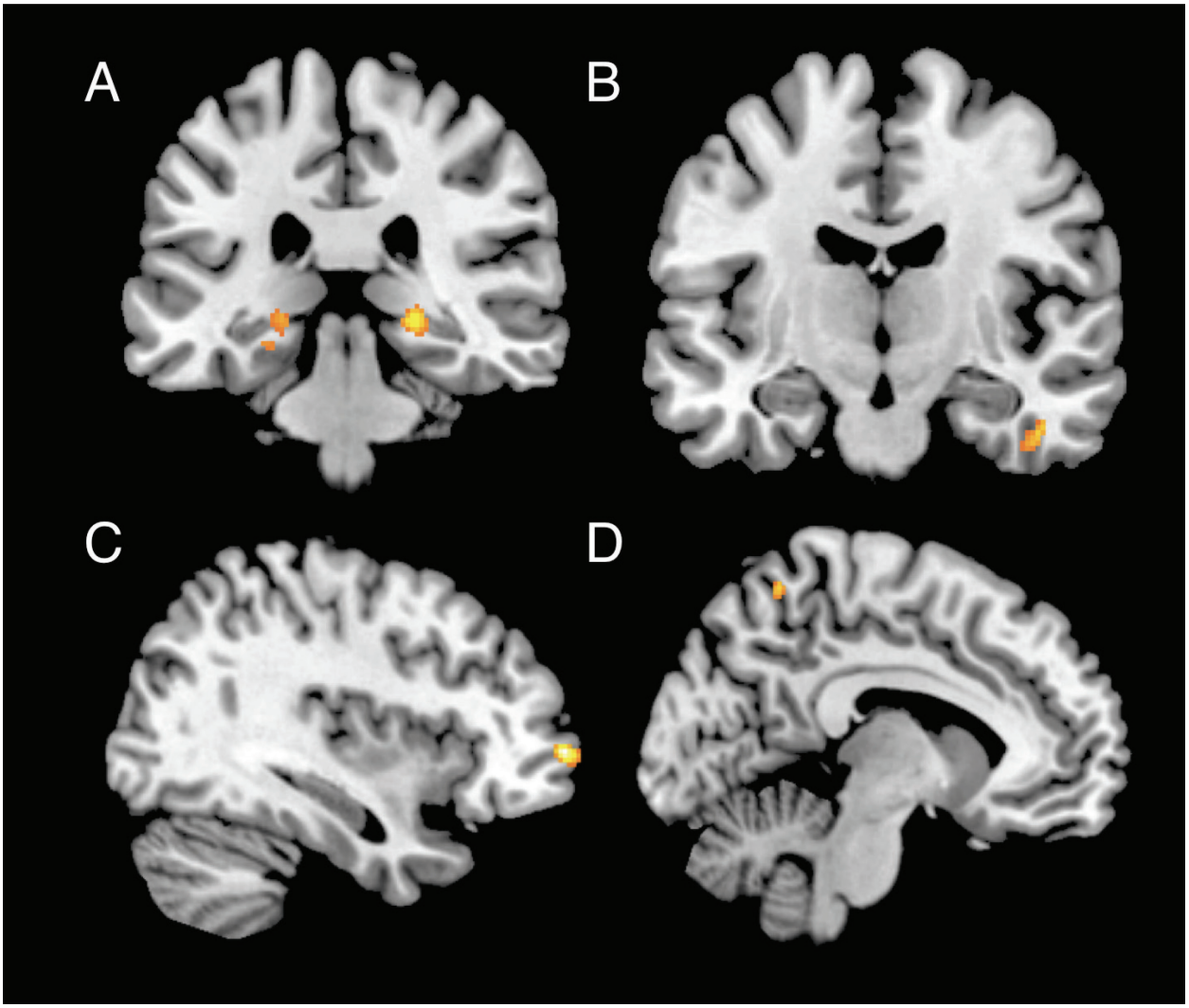

Fig. DS3 Voxel-based morphometry analysis of GM differences between LLD and controls. At a statistical threshold of $p < 0.001$ (uncorrected), the LLD group showed atrophy in (A) bilateral hippocampus; (B) R Fusiform gyrus; (C) R Frontal cortex; and (D) bilateral precuneus after controlling for TIV. Abbreviations: LLD - Late life depression; L - Left; R - Right; TIV - Total intracranial volume.

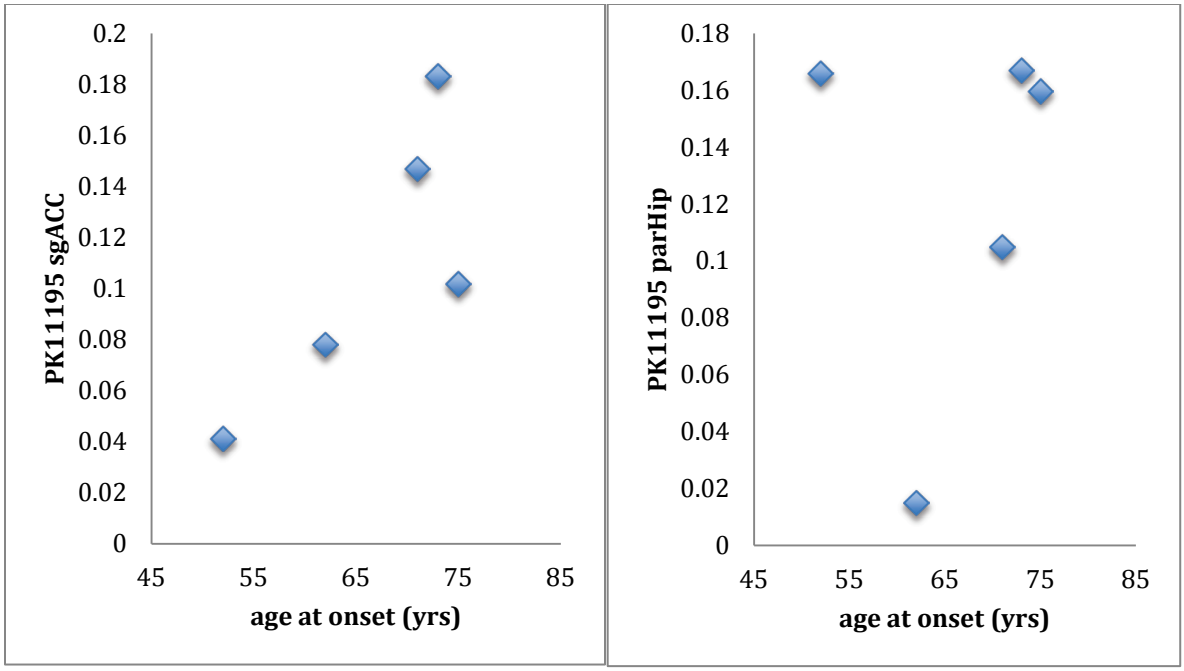

Fig DS4 Correlation between [^{11}C]PK11195 BP_{ND} and age at disease onset for five LLD subjects.

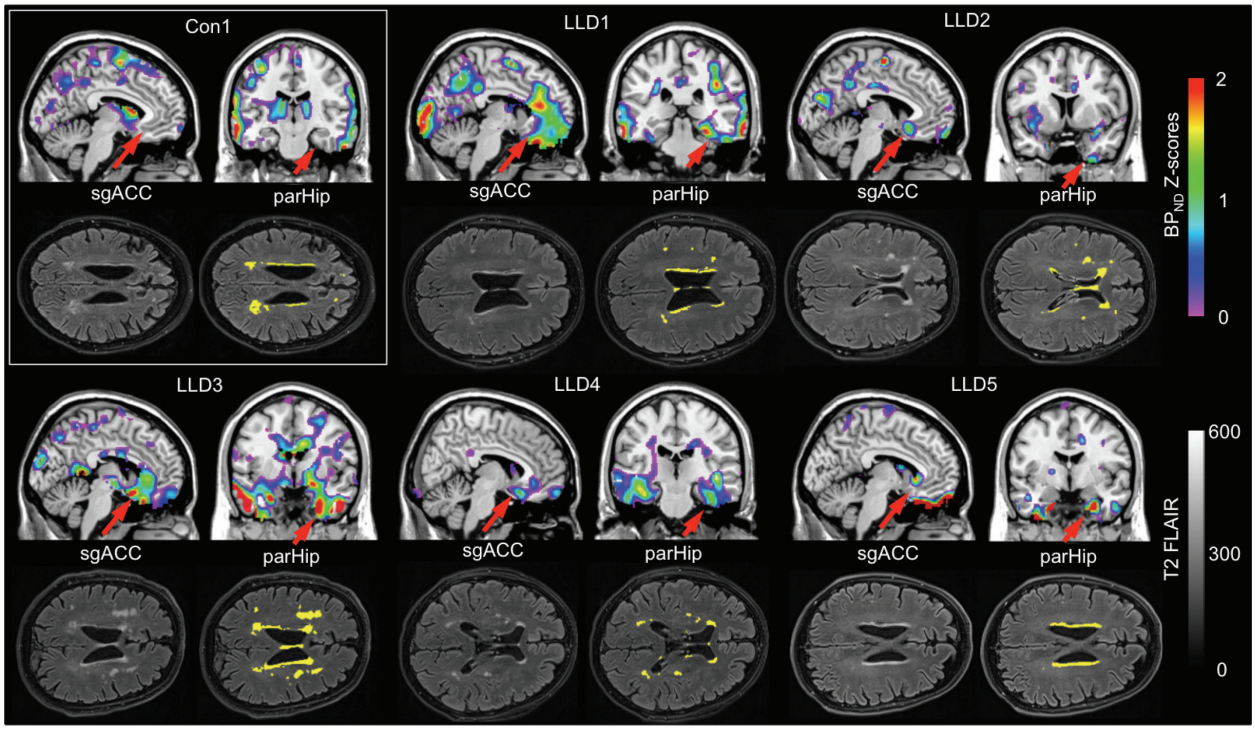

Fig DS5 Rows 1 and 3: statistical comparison of individual participant's [^{11}C]PK11195 non-displaceable binding potential (BP_{ND}) with the control group (*Z*-scores); rows 2 and 4: T_2 segmented FLAIR images with white matter lesion shown in yellow for five participants with depression. Con, control; LLD, participant with late-life depression; sgACC, subgenual anterior cingulate cortex; parHip, parahippocampus.