nature aging

ARTICLES

https://doi.org/10.1038/s43587-022-00219-7

Check for updates

Deep learning-based brain age prediction in normal aging and dementia

Jeyeon Lee[1], Brian J. Burkett[1], Hoon-Ki Min[1], Matthew L. Senjem[2], Emily S. Lundt[3], Hugo Botha[4], Jonathan Graff-Radford[4], Leland R. Barnard[4], Jeffrey L. Gunter[1], Christopher G. Schwarz[1], Kejal Kantarci[1], David S. Knopman[4], Bradley F. Boeve[4], Val J. Lowe[1], Ronald C. Petersen[4], Clifford R. Jack Jr.[1] and David T. Jones[1,4] ✉

Brain aging is accompanied by patterns of functional and structural change. Alzheimer's disease (AD), a representative neurodegenerative disease, has been linked to accelerated brain aging. Here, we developed a deep learning-based brain age prediction model using a large collection of fluorodeoxyglucose positron emission tomography and structural magnetic resonance imaging and tested how the brain age gap relates to degenerative syndromes including mild cognitive impairment, AD, frontotemporal dementia and Lewy body dementia. Occlusion analysis, performed to facilitate the interpretation of the model, revealed that the model learns an age- and modality-specific pattern of brain aging. The elevated brain age gap was highly correlated with cognitive impairment and the AD biomarker. The higher gap also showed a longitudinal predictive nature across clinical categories, including cognitively unimpaired individuals who converted to a clinical stage. However, regions generating brain age gaps were different for each diagnostic group of which the AD continuum showed similar patterns to normal aging.

The biology of aging is complex[1] and not fully understood[2]. In general, aging is characterized by the gradual accumulation of deleterious biological changes accompanying a progressive loss of function[1], although this is an oversimplification. The endeavor to better understand the biology of the aging brain is widely relevant because the impact of aging on the human brain and associated changes in cognitive function have implications for quality of life.

Brain aging entails both structural and functional changes. Structural magnetic resonance imaging (MRI) has shown that increased age is associated with reduction of gray matter volume, most prominently in the frontal lobes, insular cortex and hippocampus[3–6], increased volume of the ventricular system and intracranial cerebrospinal fluid[3,4,7] and changes in white matter microstructure[7,8]. In addition, functional imaging techniques using positron emission tomography (PET) have shown that brain aging is associated with decreased global oxygen utilization, cerebral blood flow, glucose uptake and regional changes in aerobic glycolysis[9,10]. Age-related decreased glucose utilization has been found most prominently in the frontal lobes, posterior cingulate, posterior parietal lobes[11–13] and also medial temporal regions—a critical area of pathology in dementia[14–16]. In contrast, the primary motor, occipital cortex, cerebellum and subcortical structures, including the thalamus, putamen and pallidum, are less susceptible to metabolic changes with aging[17].

Based on these findings, age prediction using brain imaging is an active area of neuroscience research[18–22]. An estimated age can be referred to as 'brain age', which may differ from the individual's chronological age[19]. Recently, growth in data availability and advancement of deep learning techniques have allowed more accurate brain age estimation in the cognitively normal population through convolutional neural network (CNN) models[21–25]. In addition, the 'brain age gap', which is the difference between brain age and chronological age, is useful as a promising, personalized biomarker of brain health[19]. On an individual basis, brain age gap measurements may also prove to have prognostic value, potentially predicting health outcomes by capturing individual differences in the interaction of aging and disease[19]. Several studies reported that an overestimation of an individual's age based on neuroimaging, measured as a large brain age gap, is associated with mortality[26], neurodegenerative diseases[27] and several other clinical conditions[19,20]. Moreover, measuring the brain age gap in cases of neurodegenerative pathology may inform our understanding of disease risk, resilience to structural/functional insults, which accumulate with aging, and the effects of diseases on the aging brain.

We aimed to develop a deep learning-based brain age prediction model using a large collection of brain structural MRI and fluorodeoxyglucose (FDG) PET scans from participants aged 26–98 years old ($n = 2{,}349$ unique individuals with 4,127 scans; cognitively unimpaired controls $n = 1{,}805$ and cognitively impaired $n = 732$). Our brain age prediction method was developed from 30–97-year-old cognitively unimpaired participants to train the healthy aging trajectories. We also studied age- and modality-specific saliency maps of the CNN model explaining which brain regions contribute most to age prediction for each age subgroup and modality type using an occlusion sensitivity analysis. We investigated the brain age gap estimation in the patient groups including mild cognitive impairment (MCI), AD, frontotemporal dementia (FTD) and dementia with Lewy bodies (DLB). The associations of brain age gap with neuropsychological tests, other imaging AD biomarkers, such as amyloid PET and tau PET, and the longitudinal predictive nature of disease progression in dementia were evaluated. A voxel-wise linear regression analysis evaluated which regional alterations contribute to higher brain age gap generation for each disease group and compared them with normal brain aging trajectories.

[1]Department of Radiology, Mayo Clinic, Rochester, MN, USA. [2]Department of Information Technology, Mayo Clinic, Rochester, MN, USA. [3]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. [4]Department of Neurology, Mayo Clinic, Rochester, MN, USA. ✉e-mail: Jones.david@mayo.edu

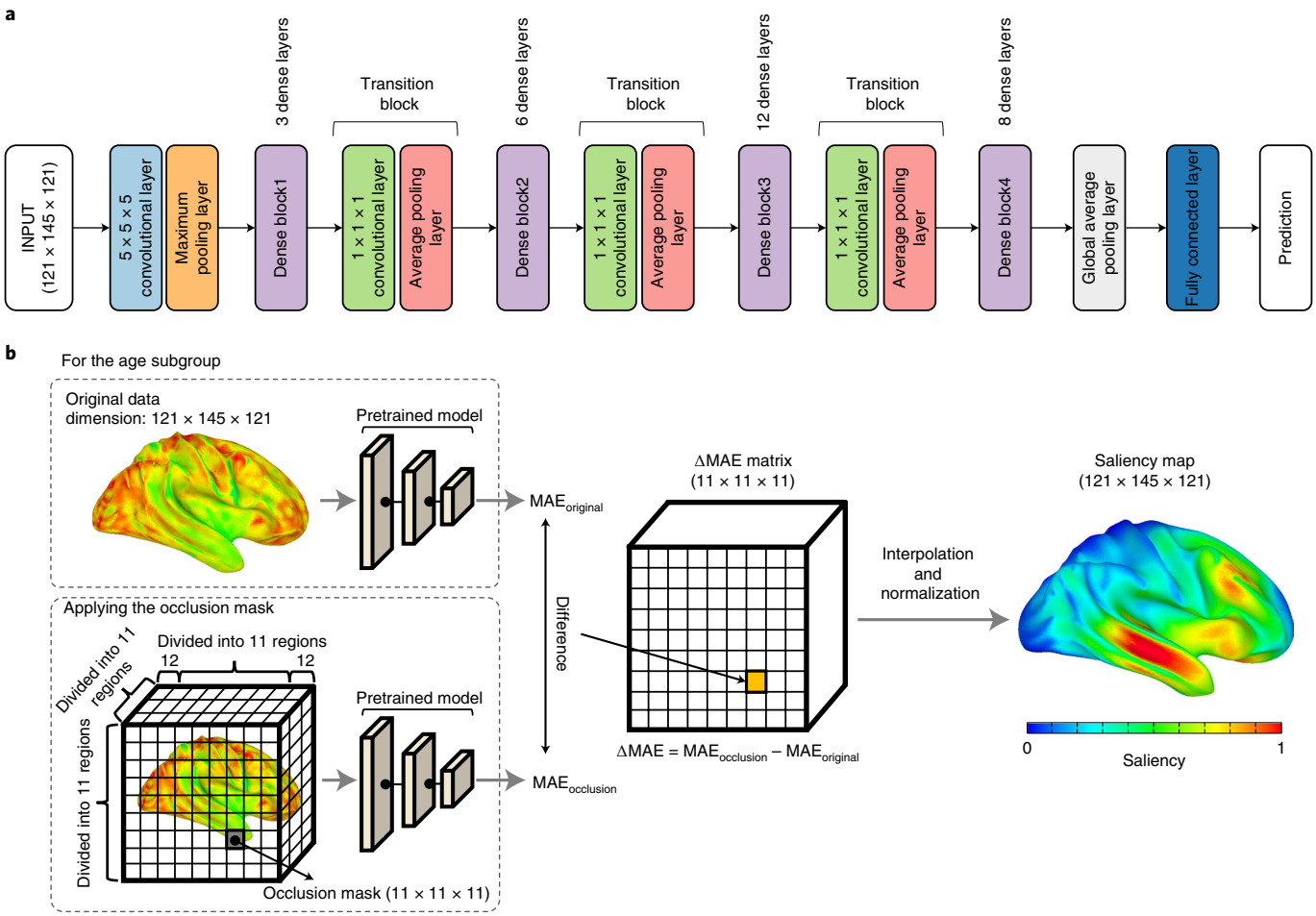


Fig. 1 | 3D-DenseNet architecture for age prediction and layout of occlusion analysis. a, Detailed architecture of the 3D-DenseNet used for age prediction. **b**, Illustration of the framework for the occlusion analysis.

Results

Brain age estimation in cognitively unimpaired participants. Our brain age prediction model based on FDG PET or MRI was trained on cognitively unimpaired participants in the Mayo dataset using a 3D-DenseNet architecture (Fig. 1a) (ref. [28]). 3D-DenseNet has good feature use efficiency and performance[28] by using dense concatenation for all subsequent layers to preserve features from the preceding layers. For training, we only utilized scans of the first time point per participant (that is, $n=1{,}805$; number of scans $n=1{,}805$) to avoid a possible overfitting problem. Then, the model's accuracy was evaluated as a mean absolute error (MAE; unit, years) with five-fold cross-validation. Fig. 2 illustrates the scatter plots of the test predictions against chronological age from a single fold. The FDG- and MRI-based model could accurately predict the chronological age of healthy adults ($R^2=0.8546$ and $\beta=0.8503$ for FDG and $R^2=0.8046$ and $\beta=0.7718$ for MRI). The overall performance measured for five folds on the test set was MAE = 3.4333 ± 0.0545 and 4.2055 ± 0.2241 for FDG and MRI, respectively (Supplementary Table 2). As shown in Fig. 2b,e, which illustrates the scatter plot of brain age gap (predicted brain age-chronological age) as a function of the corresponding chronological age, the estimation showed a tendency to be biased toward the mean age of the total cohort, resulting in a negative correlation between brain age gap and chronological age (Spearman's $r=-0.3613$ and -0.4642 for FDG and MRI, respectively). This phenomenon is well known to be associated with regression dilution[29], model regularization and a non-Gaussian age distribution[30]. We used a linear bias correction method[30] to correct age bias. After the correction, we observed that the correlation between the corrected brain age gap and chronological age decreased and the MAE also decreased (Fig. 2c,f). The overall performance after bias correction for five folds was MAE = 3.0755 ± 0.1401 and 3.4868 ± 0.1631 for FDG and MRI, respectively (Supplementary Table 2).

To assess whether the trained model presents a dataset-specific bias, the model trained with the Mayo dataset was applied to an independent cohort, the Alzheimer's Disease Neuroimaging Initiative (ADNI; https://adni.loni.usc.edu/) dataset (cognitively unimpaired, $n=330$; number of scans $n=454$). We obtained a comparable result that is not statistically different from the Mayo result (MAE = 3.1398 ± 0.2013 for FDG and MAE = 3.5101 ± 0.2270 for MRI; $P=0.58$ and $P=0.84$ for FDG and MRI, respectively; Holm-Šídák test), suggesting that the models were generalizable to the independent dataset (Extended Data Fig. 1a–f and Supplementary Table 2). In addition, we also trained a model by blending the Mayo and ADNI datasets together (Extended Data Fig. 1g–l). In this trial, the overall performance of age prediction was significantly better than using the Mayo dataset only (MAE = 2.7383 ± 0.1091 for FDG and MAE = 3.1029 ± 0.2107 for MRI; $P=0.01$ and $P=0.005$ for FDG and MRI, respectively, Holm-Šídák test; Extended Data Fig. 1m and Supplementary Table 2). The model's performance was also compared with two different architectures: 3D-ResNet and simple, fully convolutional network (SFCN)[31]. DenseNet and ResNet showed comparable performances ($P>0.05$, two-sample Student's t-test), while the performance of SFCN was poorer than DenseNet in both modalities ($P<0.001$, two-sample Student's t-test; Supplementary Table 3).

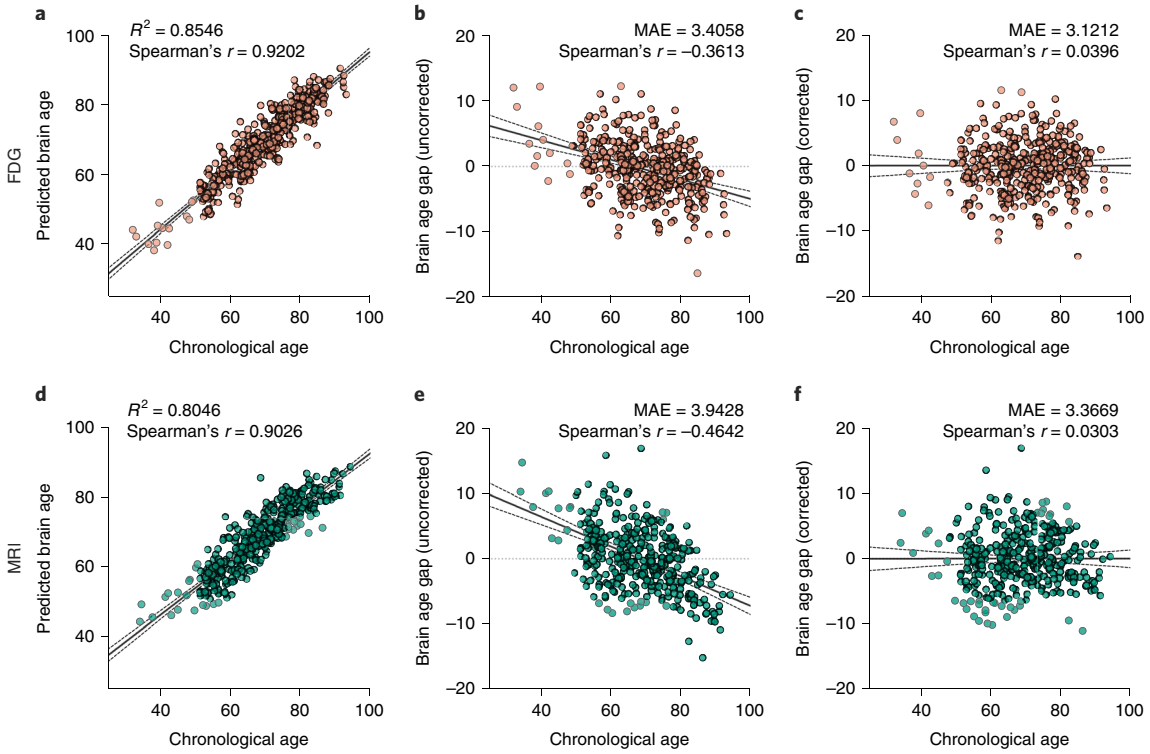


Fig. 2 | Brain age predictions on cognitively unimpaired participants. a–c, FDG-based brain age prediction result for the test set of the representative fold. **a**, Regression plot showing chronological age versus predicted brain age. **b**, Uncorrected brain age gap. **c**, Brain age gap after bias correction. **d–f**, MRI-based brain age prediction result for the test set of the representative fold. **d**, Regression plot showing chronological age versus predicted brain age. **e**, Uncorrected brain age gap. **f**, Brain age gap after bias correction. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively.

Given the longitudinal nature of our dataset, many participants had serial scans that were acquired at different time points (mean interscan interval = 2.65 ± 1.14 years for the cognitively unimpaired cohort). Although the images at each time point could be considered different and independent data because the interscan interval was long enough to allow for some changes in the acquired images, the repeated scans still had high similarity to each other. Thus, we aimed to explore whether these serial images were different enough to serve as independent data points for machine learning applications or whether this induces model overfitting and bias, thus hurting the generalizability of the model. To interrogate the possibility that within-participant variability affects the model's performance, the prediction accuracy of several data split strategies was compared (as detailed in the Methods). As expected, we observed that the overlap of the same participants between the training and validation or test datasets (assigning at least one scan of a participant to the training while assigning a different scan of the same participant to the validation or test dataset) significantly affected the accuracy of age estimation (two-sample Student's *t*-test with option 1; $P < 0.001$ for validation; test MAEs in option 2 and validation MAE in option 3; Supplementary Table 4). This pattern was similar for both FDG and MRI inputs. Meanwhile, including multiple scans for each participant showed minimal differences on the model's performance (two-sample Student's *t*-test; option 4 and option 5; Supplementary Table 4).

Saliency map of brain age prediction model. For interpretability of the trained models, saliency maps were estimated through occlusion sensitivity analysis: a portion in the input space was occluded with a mask ($11 \times 11 \times 11$) by setting these voxels to zero; their relevance in the decisions was estimated indirectly by calculating the change of MAE ($MAE_{occlusion} - MAE_{original}$; Fig. 1b). Saliency patters were age- and modality-specific (Fig. 3 and Extended Data Fig. 2). For FDG, the overall posterior region with a peak at the posterior cingulate cortex (PCC) had a higher contribution for age prediction in the younger group (30–40 and 40–50 years). For the 50–60, 60–70 and 70–80 years of age groups, the inferior frontal regions including the orbitofrontal, gyrus rectus and middle frontal regions showed a higher contribution than other areas. A global contribution with the peak around the inferior frontal cortex, basal ganglia, inferior temporal cortex and pons was also important for age prediction in the older groups (80–90 and 90–100 years). For MRI, the insular cortex contributed most to age prediction in the younger groups (30–40 and 40–50 years). From 50 to 60 years, the ventricular boundary showed a higher contribution. The cerebellomedullary cistern showed the highest saliency in the older groups (80–90 and 90–100 years). The coordinates of peak saliency found in each age range are summarized in Supplementary Table 5.

Brain age gap estimation in patient groups. The brain age gap of 4 clinical diagnosis groups (MCI, $n = 480$, number of scans, $n = 666$; AD, $n = 215$, number of scans, $n = 372$; FTD, $n = 45$, number of scans, $n = 69$; DLB, $n = 86$, number of scans, $n = 141$) was estimated using the model trained with normative cohorts. Brain age was corrected using the same coefficients used for bias correction of cognitively unimpaired individuals (Fig. 2). As expected, the brain age gap of all patient groups was significantly higher than that of the cognitively unimpaired group for both modalities ($P < 0.001$, Holm-Šídák test; Fig. 4a,c). Interestingly, the predicted brain age gap had a negative correlation with chronological age, that is, younger patients had a higher gap (Fig. 4b,d). The mean brain age gap of FTD, a relatively early onset process, was higher than that of other groups, followed by AD, DLB and MCI.

As shown in Fig. 4e, the FDG- and MRI-based brain age gaps showed significant correlation with each other ($P < 0.001$, Pearson's

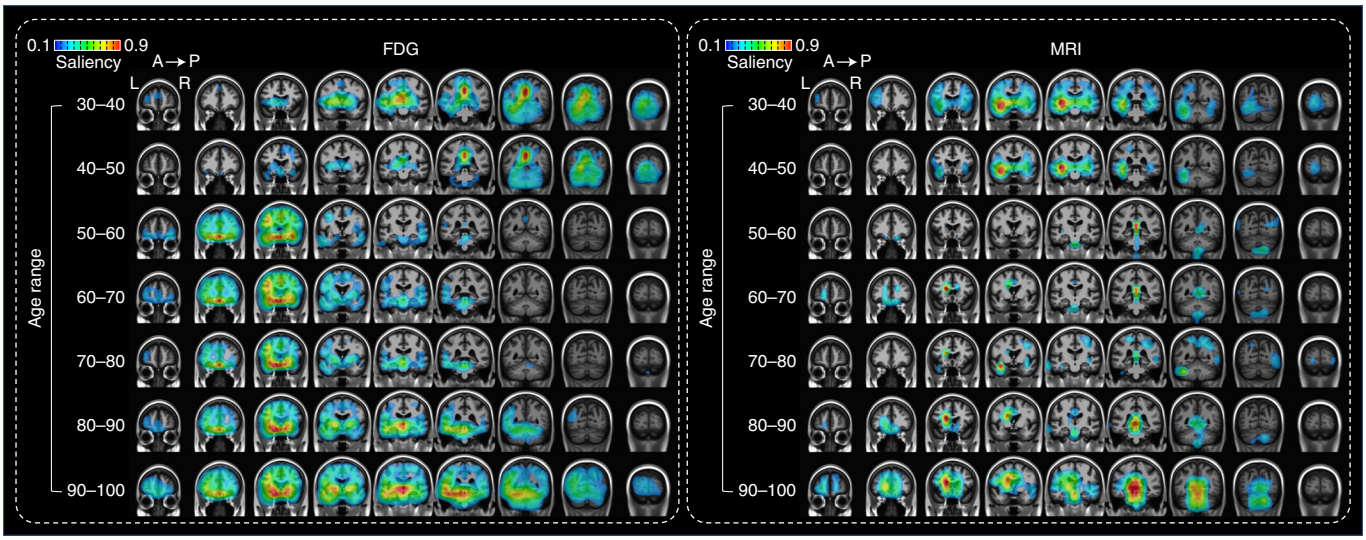


Fig. 3 | Visualization of saliency maps shown on coronal slices. Saliency maps were computed using occlusion sensitivity analysis for each age group. Higher saliency represents the importance of a region in brain age estimation. Left: Saliency maps for the FDG-based model. Right: saliency maps for the MRI-based model.

correlation coefficient) in every group. Interestingly, the disease groups tended to have a higher correlation and slope than the cognitively unimpaired groups (Pearson's correlation coefficient: 0.5819, 0.7163, 0.7974, 0.8491 and 0.6925; slope of fitted line: 0.6624, 0.7080, 0.8102, 0.8132 and 0.8126 for cognitively unimpaired, MCI, AD, FTD and DLB, respectively). The MCI, AD and FTD groups showed a significant comparison to the cognitively unimpaired groups ($P<0.001$, z-test after Fisher's r-to-z transformation; Extended Data Fig. 3).

The same models were then applied to the disease groups in the ADNI cohort (MCI, $n=647$, number of scans, $n=885$; AD, n = 255, number of scans, $n=283$) and a similar result was observed (Extended Data Fig. 4). In both modalities, the brain age gap for the MCI and AD groups was significantly higher than that of the cognitively unimpaired groups ($P<0.001$, Holm-Šídák test; Extended Data Fig. 4a,c). The correlation coefficient between the FDG- and MRI-based brain age gaps of AD was significantly higher than that of the cognitively unimpaired groups ($P<0.001$, z-test after Fisher's r-to-z transformation; Extended Data Fig. 4e).

Associations of brain age gap with demographics and AD biomarkers. A high brain age gap is linked to high cognitive impairment[19,20,25]. In light of this, the association on the corrected brain age gap of disease groups with three neuropsychological test scores was tested, including the Clinical Dementia Rating sum of boxes (CDR-SB)[32], Short Test of Mental Status (STMS)[33] and Mini-Mental State Examination (MMSE)[34]. As expected, both brain age gaps showed significant correlations ($P<0.001$, Pearson's correlation for FDG, $r=-0.3870$, -0.3762, 0.3886; for MRI, $r=-0.3612$, -0.3523, 0.3705, for MMSE, STMS and CDR-SB, respectively; Extended Data Fig. 5 and Supplementary Table 6).

Then, we sought to examine the association of brain age gap with neuroimaging AD biomarkers (Fig. 5). AD is characterized by pathological aggregation of amyloid beta and neurofibrillary tangles that can be captured by Pittsburgh Compound B (PiB) PET and tau PET, respectively. For PiB PET, only the MCI group reached statistical significance in FDG and MRI; however, the correlation coefficient was marginal and there was no obvious pattern of association in distribution (Pearson's correlation; Fig. 5a,c). However, tau PET showed a significant correlation with brain age gap in the MCI and AD groups but not FTD or DLB (Pearson's correlation; Fig. 5b,d). In particular, the AD group showed a higher correlation ($r=0.5110$ for FDG and $r=0.6648$ for MRI). The same pattern was also observed in the ADNI dataset (Extended Data Fig. 6). Only tau PET showed a significant correlation with brain age gap, while the amyloid PET did not show an association.

We also tried to evaluate the association of sex to the age gap estimation (Extended Data Fig. 7). In cognitively unimpaired individuals, females showed a significantly lower brain age gap than males in both modalities (two-sample Student's t-test, $P<0.001$ and $P=0.001$ for FDG and MRI, respectively). This is consistent with previous findings reporting that the female brain showed a persistently lower brain age compared with the male brain[18]. The brain age gap of females was estimated significantly higher than males in the AD group (two-sample Student's t-test, $P=0.009$ and $P=0.005$ for FDG and MRI, respectively). Females also showed a significantly higher brain age gap in the DLB group but this pattern was observed only in the MRI-based model (two-sample Student's t-test, $P=0.0045$).

Longitudinal predictive nature of the brain age gap. The longitudinal relevance of the brain age gap was investigated using serial scans. For this analysis, disease progression groups based on serial time points were defined as cognitively unimpaired to cognitively unimpaired, cognitively unimpaired to MCI/AD, MCI to MCI, MCI to AD, MCI to FTD, FTD to FTD, MCI to DLB and DLB to DLB, with the second category representing the most recent diagnostic group assignment. The three patients in the cognitively unimpaired to AD category were included in the cognitively unimpaired to MCI/AD, and the cognitively unimpaired to FTD ($n=1$) and cognitively unimpaired to DLB ($n=0$) groups were excluded due to an insufficient number of participants. For the cognitively unimpaired cohort, the brain age gap was collected when each participant was assigned to the test dataset. First, we explored whether the brain age gap at earlier time points (that is, baseline) was associated with the progression of diagnosis at later time points. For this comparison, only cognitively unimpaired (that is, cognitively unimpaired to cognitively unimpaired, cognitively unimpaired to MCI/AD) and MCI groups (MCI to MCI, MCI to AD, MCI to FTD and MCI to DLB) at baseline were included and the baseline brain age gap was compared within the same baseline groups (Fig. 6a,c). The cognitively unimpaired to MCI/AD group showed a significant

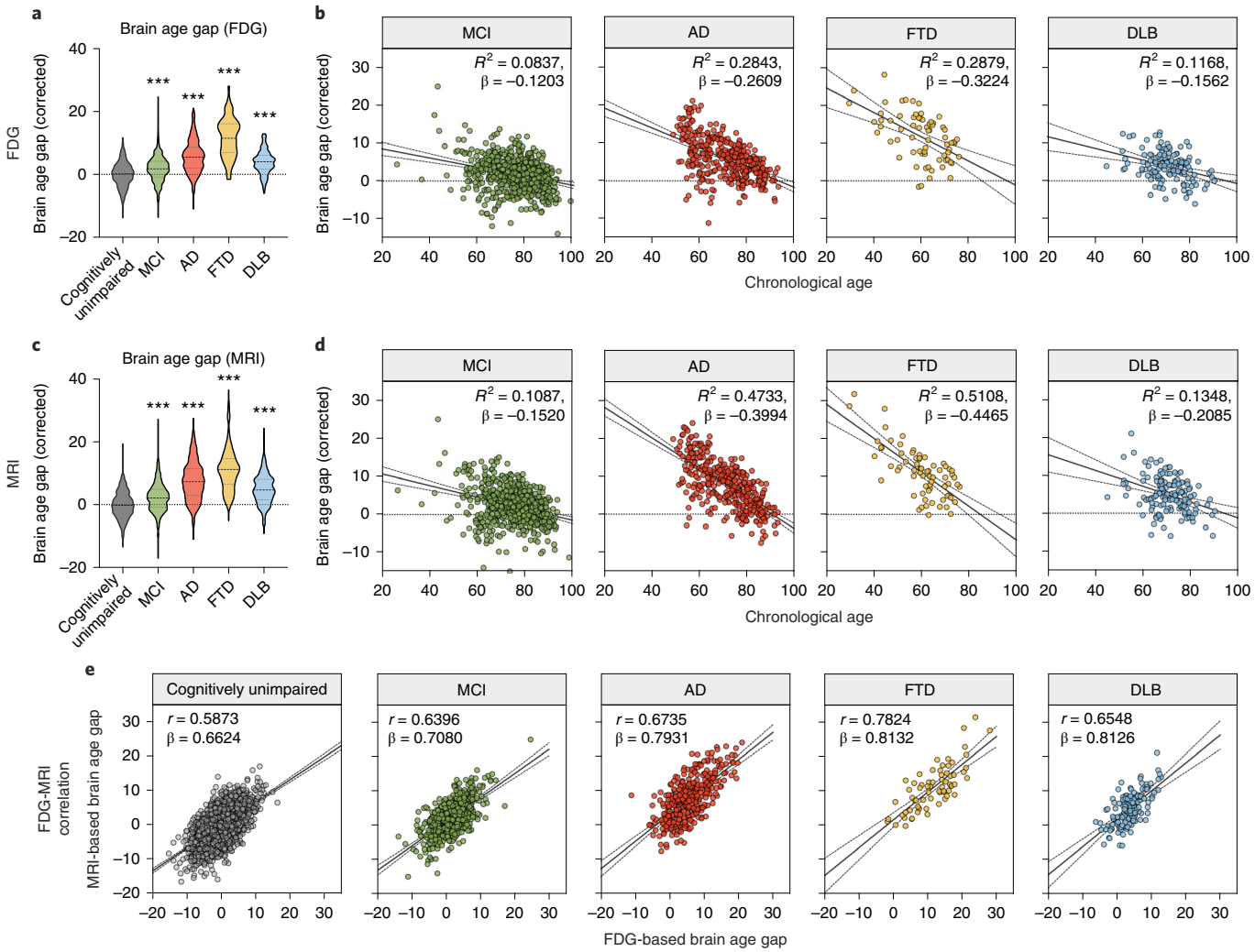


Fig. 4 | Regression plots of a corrected brain age gap as a function of chronological age for clinical diagnostic groups. a, Violin plots of corrected brain age gap for each diagnostic group. The corrected brain age gap of disease groups was compared with cognitively unimpaired individuals using a one-way ANOVA with Holm-Šídák multiple comparisons test. Exact *P* values: cognitively unimpaired versus MCI, $P=1.5\times10^{-9}$; cognitively unimpaired versus AD, $P<1\times10^{-15}$; cognitively unimpaired versus FTD, $P<1\times10^{-15}$; cognitively unimpaired versus DLB, $P<1\times10^{-15}$; ***$P<0.001$. **b**, FDG-based brain age gap estimation for MCI, AD, FTD and DLB, respectively. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. **c**, Violin plots of corrected brain age gap for each clinical diagnosis group. The corrected brain age gap of disease groups was compared with cognitively unimpaired individuals using a one-way ANOVA with Holm-Šídák multiple comparisons test. Exact *P* values: cognitively unimpaired versus MCI, $P=2.4\times10^{-11}$; cognitively unimpaired versus AD, $P<1\times10^{-15}$; cognitively unimpaired versus FTD, $P<1\times10^{-15}$; cognitively unimpaired versus DLB, $P<1\times10^{-15}$; ***$P<0.001$. **d**, MRI-based brain age gap estimation for MCI, AD, FTD and DLB, respectively. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. **e**, Relationship between FDG- and MRI-based brain age gap. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. *r* indicates Pearson's correlation coefficient.

difference of baseline brain age gap from the cognitively unimpaired to cognitively unimpaired group for both modalities ($P=0.001$ and $P<0.001$ for FDG and MRI, respectively; Holm-Šídák test). The comparison between the MCI to MCI and MCI to AD groups also reached significance in the MRI model ($P=0.005$; Holm-Šídák test), while the difference for the FDG model approached but did not meet significance ($P=0.07$; Holm-Šídák test). This observation suggests that the baseline brain age gap can predict the progression of cognitive impairment. Statistical significance of the comparison between MCI to MCI and MCI to FTD was only found in the FDG model ($P<0.001$; Holm-Šídák test) and the baseline brain age gap was not different between the MCI to MCI and MCI to DLB groups for either imaging modality. Since the probability of disease progression increases with longer time intervals between consecutive scans, the interscan interval between groups was also compared. We found that the interval of the cognitively unimpaired to cognitively unimpaired group was significantly higher than that of the cognitively unimpaired to cognitively unimpaired group to the MCI/AD groups (Extended Data Fig. 8a; Holm-Šídák test). To exclude any bias due to the difference in interscan interval, we repeated the comparison of baseline brain age gap after excluding participants with an interscan interval of >2 years (Extended Data Fig. 8b) and found that the baseline brain age gap was still predictive of disease progression at a later time point in both modalities (Holm-Šídák post hoc test; Extended Data Fig. 8c,d).

A similar result was observed in the external ADNI cohort (Extended Data Fig. 9a,c). The comparison between the cognitively unimpaired to cognitively unimpaired and cognitively unimpaired

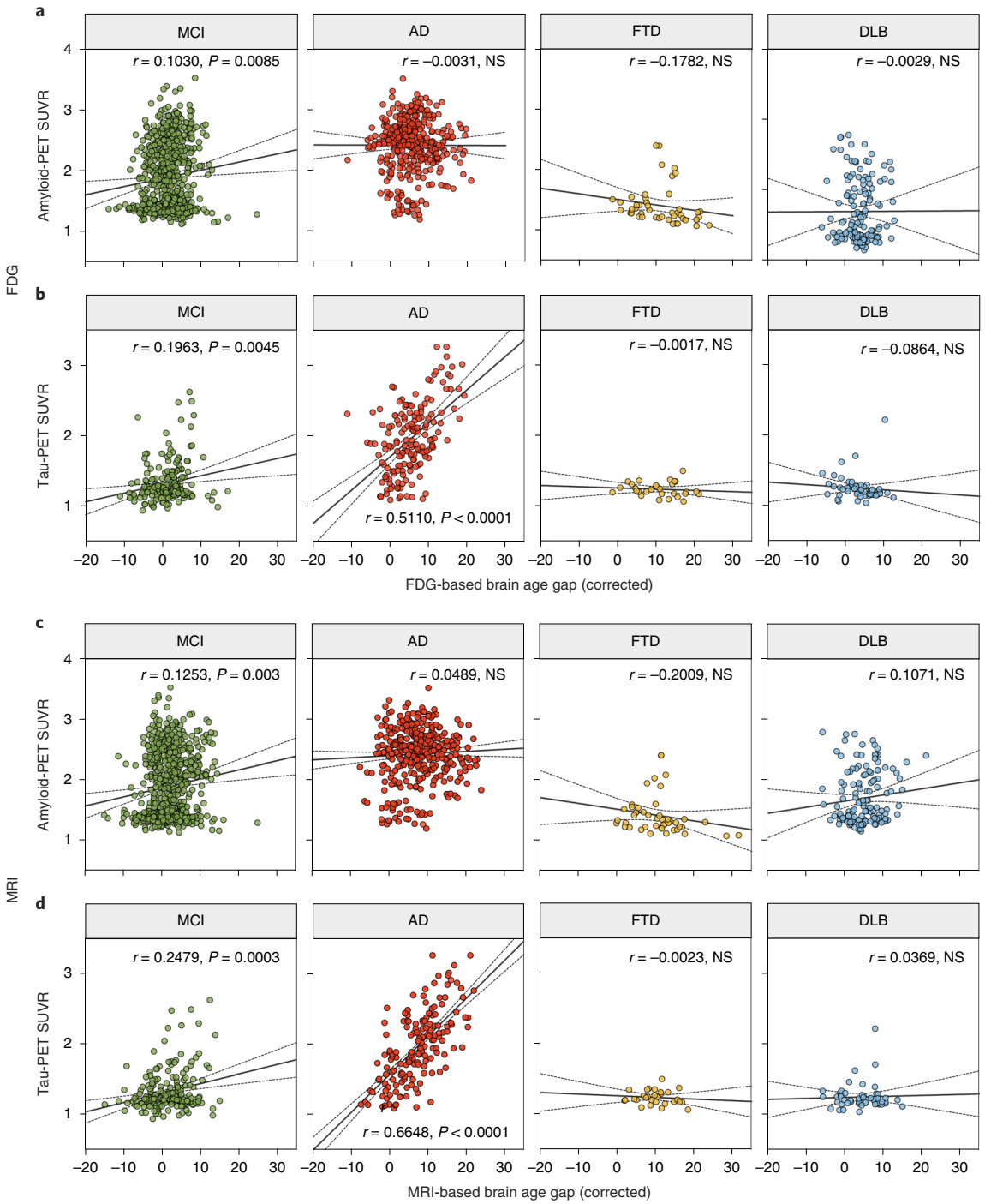


Fig. 5 | Association of brain age gap with meta-ROI PiB and Tau PET SUVR. a, Scatter plots show the relationship between FDG-based brain age gap with meta-ROI PiB PET SUVR for MCI, AD, FTD and DLB, respectively. **b**, Scatter plots of FDG-based brain age gap versus meta-ROI tau PET SUVR. **c**, Scatter plots of MRI-based brain age gap versus meta-ROI PiB PET SUVR. **d**, Scatter plots of MRI-based brain age gap versus meta-ROI Tau PET SUVR. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. *r*, Pearson's correlation coefficient; *P*, correlation test *P* value.

to MCI/AD groups was statistically significant in both models ($P=0.04$ and $P<0.001$ for FDG and MRI, respectively; two-sample Student's *t*-test). Only the MRI model showed a statistically significant difference in comparison between MCI to MCI and MCI to AD ($P=0.03$; two-sample Student's *t*-test).

Next, we looked at how the longitudinal change of brain age differed for each disease group. For this analysis, the annual rate of change in brain age gap (Δbrain age gap per year) between consecutive scans was compared between the groups. Thus, we found that in both modalities, the MCI to AD and AD to AD groups showed a significantly higher Δbrain age gap than the cognitively unimpaired to cognitively unimpaired group (Holm-Šídák test; Fig. 6b,d). Only the FDG model showed statistical significance in the FTD to FTD group ($P<0.001$, Holm-Šídák test; Fig. 6b). In the ADNI cohort, only the AD to AD group in the FDG model showed a significantly higher Δbrain age gap compared to the cognitively unimpaired

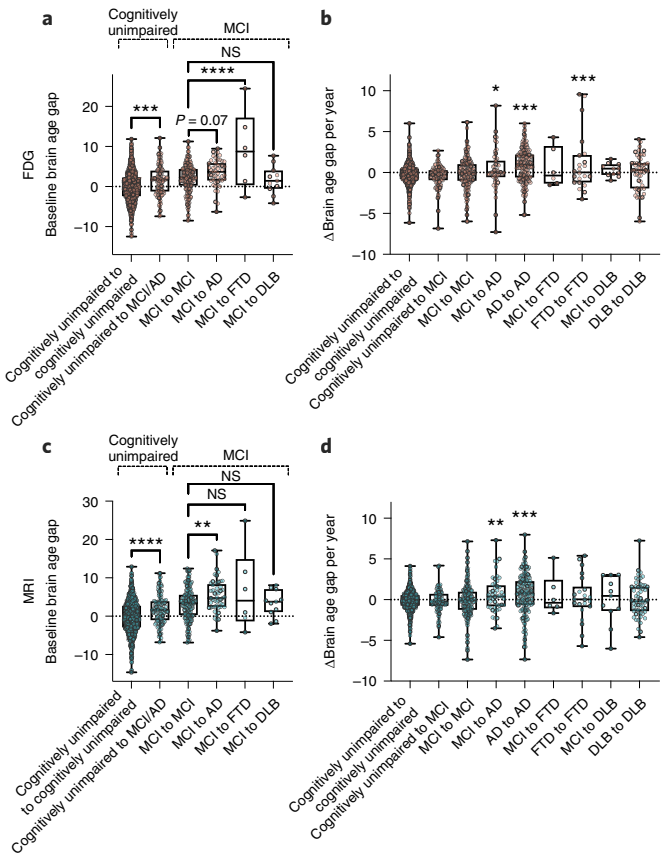


Fig. 6 | Longitudinal nature of the brain age gap. The disease progression group seen in participants at the serial scans was defined with the second category representing the most recent diagnostic group assignment. **a**, Baseline brain age gap comparison for the FDG model. A statistical test was performed within the same baseline groups using a one-way ANOVA with Holm-Šídák post hoc test. Exact *P* values: cognitively unimpaired to cognitively unimpaired versus cognitively unimpaired to MCI/AD, $P = 1.9 \times 10^{-4}$; MCI to MCI versus MCI to AD, $P = 0.07$; MCI to MCI versus MCI to FTD, $P = 2.6 \times 10^{-5}$; MCI to MCI versus MCI to DLB, $P = 0.78$. **b**, For the FDG model, the annual Δ brain age gap of each group was compared with the cognitively unimpaired to cognitively unimpaired using a one-way ANOVA with Holm-Šídák post hoc test. Exact *P* values: cognitively unimpaired to MCI/AD, $P = 0.54$; MCI to MCI, $P = 0.57$; MCI to AD, $P = 0.04$; AD to AD, $P < 1.0 \times 10^{-15}$; MCI to FTD, $P = 0.56$; FTD to FTD, $P = 3.5 \times 10^{-4}$; MCI to DLB, $P = 0.57$; DLB to DLB, $P = 0.57$. **c**, Baseline brain age gap comparison for the MRI model. A statistical test was performed within the same baseline groups using a one-way ANOVA with Holm-Šídák post hoc test. Exact *P* values: cognitively unimpaired to cognitively unimpaired versus cognitively unimpaired to MCI/AD, $P = 8.2 \times 10^{-5}$; MCI to MCI versus MCI to AD, $P = 0.005$; MCI to MCI versus MCI to FTD, $P = 0.11$; MCI to MCI versus MCI to DLB, $P = 0.87$. **d**, For the MRI model, the annual Δ brain age gap of each group was compared with the cognitively unimpaired to cognitively unimpaired using a one-way ANOVA with Holm-Šídák post hoc test. Exact *P* values: cognitively unimpaired to MCI/AD, $P = 0.99$; MCI to MCI, $P = 0.99$; MCI to AD, $P = 0.006$; AD to AD, $P < 3.5 \times 10^{-7}$; MCI to FTD, $P = 0.87$; FTD to FTD, $P = 0.87$; MCI to DLB, $P = 0.99$; DLB to DLB, $P = 0.87$. $n = 1{,}054, 104, 169, 49, 157, 6, 22, 11, 55$ for cognitively unimpaired to cognitively unimpaired, cognitively unimpaired to MCI/AD, MCI to MCI, MCI to AD, AD to AD, MCI to FTD, FTD to FTD, MCI to DLB and DLB to DLB groups, respectively. The box plots represent the minimum and maximum values (whiskers), the first and third quartile (box boundaries) and the median (internal line). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

to cognitively unimpaired group ($P < 0.001$, Holm-Šídák test; Extended Data Fig. 9b).

Brain age gap in dementia and normal aging. A voxel-wise linear regression analysis was performed using the brain age gap as a regressor to investigate which brain region alterations were related to higher brain age gap generation for each patient group. In this analysis, chronological age was specified as a nuisance covariate because it was negatively correlated with the brain age gap. The FDG- and MRI-based brain age gap showed different patterns according to the disease groups (using linear regression, false discovery rate (FDR)-corrected, $q < 0.01$; Fig. 7a,b). In FDG, the MCI and AD groups showed a negative correlation throughout the brain, meaning that global cortical hypometabolism was associated with a higher brain age gap, while the white matter region showed positive correlation (Fig. 7a and Extended Data Fig. 10). In the AD group, the frontal, temporal and parietal regions showed a stronger negative correlation. In contrast, significant hypometabolism related to the brain age gap was observed in the frontal and temporal regions in the FTD patient group. Interestingly, the occipital, precentral cortex and thalamus showed a positive correlation in the FTD group. The DLB group showed a significant negative correlation in the posterior and temporal regions. The precentral cortex and thalamus showed a positive correlation with the brain age gap. However, MRI showed a distinctly different pattern of prominent regions from FDG (Fig. 7b and Extended Data Fig. 10). In MCI and AD, sulci and white matter showed a positive correlation; regions around the gyri and ventricles showed a negative correlation with the brain age gap. In contrast, a local negative correlation around the ventricles was marginally observed for the FTD and DLB patient groups.

To compare the observed brain age gap-related changes with normal aging, a linear regression analysis was also performed for the cognitively unimpaired group using chronological age as a regressor (Fig. 7c,d). Compared to the occlusion analysis, voxel-wise regression can highlight brain regions showing statistically significant associations with normal aging. The regression analysis of age in the cognitively unimpaired group differed from the salience analysis on the same group of images, a plausible result because the model may have focused on specific features (even statistically nonsignificant features) rather than treating all input information together as a group as with the regression. Like the results for MCI and AD, a global cortical negative correlation and positive correlation in white matter were observed on FDG (Fig. 7c). A positive correlation in sulci and white matter and a negative correlation in areas around the gyri and ventricles was observed on MRI (Fig. 7d). Then, to evaluate the similarity between the brain age gap-related changes and normal aging, a voxel-wise correlation analysis between beta values was performed (Fig. 7e–h). The result showed that the beta map of each patient group was strongly correlated with that of normal aging for FDG and MRI ($P < 0.001$, Pearson's correlation). The similarity of MCI was strongest among groups and followed by the AD group in both modalities ($P < 0.001$, *z*-test after Fisher's *r*-to-*z* transformation; Fig. 7f,h). The correlation coefficients of the FTD and DLB groups were relatively lower than those of the MCI and AD groups ($P < 0.001$, *z*-test after Fisher's *r*-to-*z* transformation; Fig. 7f,h).

Discussion

We developed the 3D-DenseNet models, trained on structural or metabolic brain images, which accurately estimated an individual's brain age during normal aging. An occlusion analysis revealed anatomical regions critical to the model performance and demonstrated an age-dependent saliency pattern of brain regions, which was distinct for each input imaging modality. In cohorts with a neurological disorder, the brain age gap was larger than in cognitively unimpaired individuals and was significantly correlated with

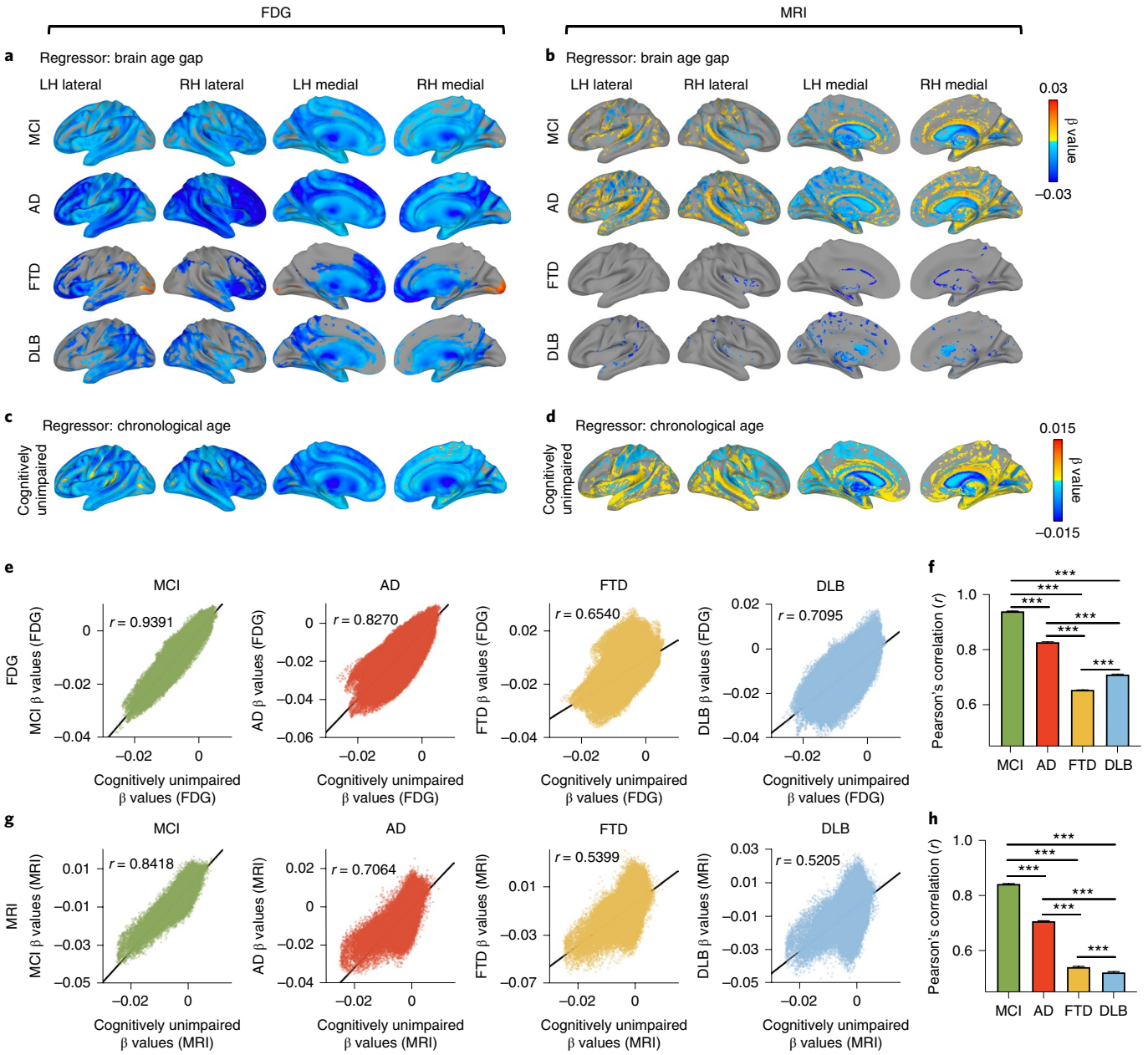


Fig. 7 | Voxel-wise linear regression analysis of the brain age gap. a,**b**, Clinical diagnosis group-specific (MCI, AD, FTD and DLB) results from voxel-wise whole-brain linear regression examining brain age gap-related changes (FDR-corrected, $q < 0.01$) for FDG and MRI, respectively. Chronological age was specified as a nuisance covariance. **c**,**d**, For cognitively unimpaired individuals, voxel-wise linear regression analysis was performed using the chronological age as a regressor to show the age-related change for FDG (**c**) and MRI (**d**), respectively. **e**, Voxel-wis**e** correlation of beta value between clinical diagnosis group (vertical axis) and cognitively unimpaired group (horizontal axis) for the FDG model. **f**, Comparisons of Pearson's correlation coefficients from **e**. The error bars indicate 95% confidence intervals (CIs) of the Pearson's correlation coefficient. Pearson's r (95% CI) = 0.9391 (0.9386–0.9396), 0.827 (0.8257–0.8283), 0.654 (0.6517–0.6563) and 0.7095 (0.7114–0.7075) for MCI, AD, FTD and DLB, respectively. All exact P values are $P < 1.0 \times 10^{-15}$. ***$P < 0.001$, two-sided z-test after Fisher's r-to-z transformation. **g**, Voxel-wise correlation of beta value between clinical diagnosis group (vertical axis) and cognitively unimpaired group (horizontal axis) for the MRI model. **h**, Comparisons of Pearson's correlation coefficients from **g**. The error bars indicate the 95% CIs of the Pearson's correlation coefficient. Pearson's r (95% CI) = 0.8418 (0.8407–0.8430), 0.7064 (0.7043–0.7084), 0.5399 (0.5370–0.5427) and 0.5205 (0.5176–0.5235) for MCI, AD, FTD and DLB, respectively. All exact P values are $P < 1.0 \times 10^{-15}$. ***$P < 0.001$, two-sided z-test after Fisher's r-to-z transformation.

the cognitive score and AD neuroimaging biomarker. Additionally, the brain age gap measure at baseline predicted the progression of cognitive impairment at a later time point. Anatomical regions with the greatest association with the brain age gap, identified from the voxel-wise linear regression analysis, were different for each diagnostic group. The results for the AD continuum, MCI and AD, showed close correlations to normal aging compared to FTD or DLB.

Most previous brain age studies were based on structural MRI[21–26,31]. To our knowledge, only 1 prior study utilized FDG PET[18] but that study was based on a non-deep learning method and utilized a substantially smaller cohort ($n = 205$). A limited number of

structural MRI-based studies reported explanation maps of the CNN model[21,22,24]. The structural and functional changes contributing to precise age prediction in the deep learning approach are to be fully elucidated. The longitudinal predictive nature of the brain age gap has not yet been explored in a preclinical group. Furthermore, there is a dearth of knowledge regarding which brain alterations and specific regional changes are associated with higher brain age gaps in patients and the relationship of expected biological senescence and pathological processes.

Our model precisely estimated an individual's chronological age based on structural and metabolic neuroimaging data. Interestingly, FDG-based brain age prediction was slightly better than the MRI-based model (Fig. 2 and Supplementary Table 2), suggesting that metabolic data may be more sensitive for tracking normal brain aging trajectories. One consideration is that metabolic changes detectable on PET may precede structural changes observed in AD[35], although this has not been characterized in cognitively unimpaired individuals. Also, our FDG-based model partially incorporated structural information since the spatial normalization to template space for the FDG scan was performed using the individual's MR images, meaning that the brain age prediction model using FDG has the benefit of both functional and structural information. The FDG images are also affected by structural changes via partial volume effects. Alternatively, the decreased performance of the model using MRI relative to FDG could be a consequence of regional heterogeneity in age-related structural changes in the brain[36].

Occlusion analysis revealed regions important for age estimation and showed a distinct age-specific saliency pattern according to the input imaging modality (Fig. 3 and Extended Data Fig. 2). In the FDG-based model, a transition of posterior to anterior structures with increased age was observed. The posterior structures, especially the PCC, contributed most in the younger age groups, whereas anterior structures including the frontotemporal lobes were more critical in the older age groups. Glucose metabolism decline in the PCC with age has been reported[11] and amyloid deposition and reduced glucose metabolism in the PCC has been implicated in early AD[37]. In older adults, FDG activity in the frontal regions was more salient; decline of frontal metabolism in normal aging was consistently reported across several studies[14,38]. The MRI-based model's saliency map demonstrated different critical regions compared to the FDG analysis. For the younger age groups, the insula was identified as the most critical region, a region that undergoes gray matter volume loss with normal aging[39]. Additionally, the medial temporal lobe was identified as an area with high saliency in the MRIs of younger, 30–50-year-old individuals, regions of previously described volume loss with aging as well as AD[40]. Preservation of brain parenchyma in the insula and medial temporal lobe of younger individuals may have been a reliable feature for MRI-based age prediction. For the older age groups, the cerebellomedullary cistern and peripheral boundaries of the ventricles were critical. This may reflect reliance of the model on the typical enlargement of the cerebrospinal fluid (CSF) spaces that occurs with age[3,4,7,41]. Interestingly, the saliency maps did not show a prominent contribution of cortical regions for age estimation, which we expected to find due to the typical age-dependent decrease in cortical volume seen on MRI[39,41]. We speculate that cortical changes with age may be too heterogenous to serve as the most reliable salient feature for the age prediction model. Changes in white matter signal characteristics are also a well-known phenomenon of aging[42]. No contribution of white matter was found with our occlusion analysis, which might be a consequence of white matter intensity normalization performed on MRI.

Interestingly, the estimated brain age gap was negatively correlated with chronological age for both MRI and FDG and was close to zero in the older age groups (Fig. 4), suggesting that the model cannot distinguish normal from diseased brain at a similar older age. Alternatively, attrition could explain the negative association between chronological age and brain age gap because individuals with a more diseased brain (that is, higher brain age gap) are not likely to survive to older ages. The brain age gap of MCI and AD showed a significant association with tau PET but not amyloid PET (Fig. 5 and Extended Data Fig. 6). Tau is well known to be more closely related to AD severity than amyloid level[43]. In both preclinical AD and AD dementia, tau radiotracer uptake and cortical thickness are correlated with decreased cognitive task performance to a greater degree than amyloid beta radiotracer uptake[43]. Furthermore, the brain age gap estimation is capable of predicting disease progression even in the preclinical stage in a longitudinal design (Fig. 6). One prior study reported that a higher brain age gap was found in progressive MCI changing diagnosis from MCI at baseline to AD at follow-up compared to the stable MCI group[44]. Franke and Gaser[44] showed that brain aging accelerates more in progressive MCI and AD groups than cognitively unimpaired individuals and stable MCI[44].

A strong correlation was observed between FDG- and MRI-based brain age gap in the cognitively unimpaired and neurodegenerative disease groups. This suggests that the metabolic changes of normal aging, as well as disease progression, are concurrent with structural changes, with regard to factors that impact the performance of the age prediction model. The correlation between FDG- and MRI-based brain age gap was mildly stronger in the disease groups ($r = 0.6548–0.7824$) than in the cognitively unimpaired cohort ($r = 0.5873$). The structural changes or atrophy in neurodegenerative pathology accompanying hypometabolism, to a greater extent than with normal aging, is one plausible explanation for the increased correlation in the disease groups. Alternatively, brain hypometabolism, which occurs in specific patterns for different categories of neurodegenerative pathology[45], may correlate more closely with structural or volumetric changes for specific neurodegenerative disease cohorts than in normal aging.

In FTD, the frontal and anterior temporal regions showed a negative correlation with the brain age gap, regions with characteristic hypometabolism in FTD;[46,47] a positive correlation was observed in the occipital lobe, a region typically without hypometabolism in FTD[46,47]. Castelnovo et al.[48] reported that some cases with FTD showed occipital hypermetabolism[48]. In DLB, the temporal, parietal and occipital regions were negatively correlated with the brain age gap, regions of hypometabolism frequently observed in DLB[46]. Correlation between the occipital lobe and primary visual cortex in the DLB group is notable because occipital/primary visual cortex hypometabolism is characteristic of DLB from other neurodegenerative processes such as AD[46,49]. The ability of the metabolic signature to distinguish DLB from AD is unique and an important component of the clinical utility of FDG PET[50] since abnormal amyloid PET, which is a defining hallmark of AD, is commonly present in DLB due to the phenomenon of co-occurring pathologies with advancing age[49]. The ventricle and boundaries of the brain parenchyma with the CSF space were correlated with MCI and AD in MRI. For FTD and DLB, the ventricular boundary was correlated with the brain age gap, although no correlation was seen at the CSF and cortical region. Periventricular borders with the CSF may reflect areas of white matter volume loss and the gyral/sulcal interface, both also occurring with normal aging[3,4,7]. The decreased correlation found in the FTD and DLB groups relative to the MCI and AD groups could be due to the smaller sample size in the FTD and DLB cohorts.

Brain age estimation has potential as a useful neuroscientific and prognostic clinical tool, although the conceptual paradigm underlying a 'brain age' has attracted some criticism and debate of which deviations are due to the result of specific pathological processes rather than an acceleration of normal biological senescence[19]. Using both functional and structural neuroimaging, we demonstrate that brain age gap-associated changes in MCI and AD have a stronger similarity to normal aging than those of FTD and DLB (Fig. 7). Brain

age gap estimation in MCI and AD using our model may reflect a process of accelerated aging versus FTD and DLB in which the brain age gap may represent specific regional pathology. However, pathological entities and normal aging changes cannot be easily dissociated because aging results from cumulative biological damage[51], which suggests that biological aging and disease are intrinsically linked[19]. Relevant to this conceptual framework, we showed that MCI correlates with normal aging more than AD, suggesting that the mild disease condition is more similar to biological aging. With greater severity in the AD continuum, the pattern of changes was more pathology-specific. In this sense, brain age estimation may more provide greater insights, informing our understanding of the relationship of the aging process to degenerative pathology in a broader sense. If dementia reflects a continuum of the underlying changes in brain structure and metabolism to which all individuals are inevitably susceptible at various rates, brain age prediction based on neuroimaging may yield a better understanding of different brain aging phenotypes. Alternatively, if types of dementia represent entities with distinctly different mechanisms than normal aging, markers of brain age may still prove useful in identifying individuals at greater risk for developing these conditions[52,53].

This study has some notable limitations. In the occlusion analysis, left hemispheric dominance was observed in the contribution to brain age prediction, which was not explainable by post hoc analysis. The occlusion-based method focuses more on the most dominant regions compared to other interpretation methods[54]. In this study, we only tested neurodegenerative pathology without evaluating any chronic systemic medical diseases and vascular diseases that may have different patterns of brain aging. Previous studies[21,25] trained CNNs on larger MRI samples (approximately 10,000) by aggregating public datasets and were able to greatly improve performance. We did not include cohorts from additional public datasets in which FDG was not available because we aimed to compare MRI- and FDG-based models to test how the two modalities behave differently in age prediction tasks for cognitively unimpaired and dementia groups. Although we acquired successful prediction accuracies, increasing the sample size for training could further improve the model's performance. Our model presented reasonable performance; however, excluding those voxels absent of brain parenchyma could reduce the parameters and further increase training efficiency.

In summary, we showed that a 3D-DenseNet model generates accurate age prediction for cognitively unimpaired individuals, with slightly more robust performance using an FDG PET input than MRI. Brain age prediction using PET imaging, which reflects metabolic function, may present a distinct assessment of brain health from the structural information evaluated on MRI. The brain age gap from MRI or FDG data is increased in multiple types of dementia compared to cognitively unimpaired individuals; therefore, it may prove to be a useful composite biomarker to identify increased risk for pathology or as a marker of disease severity.

Methods

Dataset. A large number of participants (Table 1) ranging in age from 26 to 98 years old were included ($n=2,349$, number of scans = 4,127) who had both MRI and FDG PET from the Mayo Clinic Study of Aging (MCSA) or the Alzheimer's Disease Research Center (ADRC) study (Table 1). All participants or designees provided written informed consent with the approval of Mayo Clinic and Olmsted Medical Center institutional review boards. As described previously, the Mayo Clinic Rochester ADRC is a longitudinal cohort study that enrolls participants from the clinical practice at the Mayo Clinic in Rochester[55]. The MCSA is a population-based study of cognitive aging among Olmsted County residents[56]. Enrolled participants are adjudicated to be clinically normal or cognitively impaired by a consensus panel consisting of study coordinators, neuropsychologists and behavioral neurologists. Methods for defining clinical unimpairement, MCI and dementia in both of these studies conform to standards in the field[57–59]. For this analysis, participants were assigned into 6 clinical subgroups based on clinical diagnosis according to consensus criteria[49,60] including cognitively unimpaired ($n=1,805$, number of scans = 2,879), MCI ($n=480$, number of scans = 666), AD ($n=215$, number of scans = 372), FTD ($n=45$, number of scans = 69) and DLB ($n=86$, number of scans = 141).

For the CNN model training, only data from cognitively unimpaired individuals were utilized. Some participants also underwent amyloid PET scanning with PiB (number of scans = 2,508) and tau PET scans with flortaucipir (number of scans = 608). Some participants had CDR-SB[32], STMS[33] and MMSE[34] available (number of scans = 1,522, 1,491 and 1,587, respectively). All cognitive tests were administered by experienced psychometrists and supervised by board-certified clinical neuropsychologists. To examine whether the trained model presented a dataset-specific bias, we also utilized the ADNI dataset ($n=1,150$, number of scans = 1,622; Supplementary Table 1). The ADNI dataset included cognitively impaired participants ($n=330$, number of scans = 454) and participants with MCI ($n=647$, number of scans = 885) and dementia ($n=255$, number of scans = 283). Some participants from the ADNI dataset also underwent amyloid PET scanning with AV45 (number of scans = 1,464) and tau PET scans with flortaucipir (number of scans = 283).

Image processing. T1-weighted MRI scans were acquired using 3T scanners. FDG PET imaging was performed with ^{18}F-fluorodeoxyglucose, amyloid PET with PiB[61] and tau PET with ^{18}F-flortaucipir (AV-1451) (ref. [62]). FDG PET images were acquired from 30–40 min, PiB PET from 40–60 min and tau PET from 80–100 min after injection. Computed tomography was obtained for attenuation correction. PET images were analyzed with our in-house fully automated image processing pipeline[63]. Briefly, PET scans were co-registered to the corresponding MRI for each participant within each time point and subsequently warped to the Mayo Clinic Adult Lifespan Template (MCALT) space[64] (https://www.nitrc.org/projects/mcalt/) using the warps from SPM12 unified segmentation[65]. The corresponding MRI was corrected for intensity inhomogeneity and segmented using MCALT tissue priors and segmentation parameters. The FDG PET standardized uptake value ratio (SUVR) was calculated by dividing the median of uptake in the pons and the SUVR images were used for input data to the CNN model. The amyloid and tau PET SUVR were calculated by dividing the median uptake in the cerebellar crus gray matter[66]. A meta-region of interest (ROI) PiB PET SUVR was derived from the average of the median SUVR in the prefrontal, orbitofrontal, parietal, temporal, anterior cingulate and posterior cingulate/precuneus regions[66]. A meta-ROI tau PET SUVR was formed from the average of the median uptake in the amygdala, entorhinal cortex, fusiform, parahippocampal and inferior temporal and middle temporal gyri[66]. For each MRI volume, the intensities of the voxels were normalized by dividing the mean intensity derived from the individualized white matter mask after spatial normalization. Then, this image was used for the input data to the CNN model[67].

3D-DenseNet architecture and training. A modified 3D-DenseNet model[28] was trained on FDG PET or MRI scans of the cognitively unimpaired cohort (Fig. 1a). For the training, we only utilized scans of the first time point (number of scans = 1,805) to avoid data leakage between the training and validation/test datasets. Experimental tests measuring how an overlap of participants among training, validation and test datasets affected the model's results were performed separately (see the Dataset split experiment section). A schematic of the 3D-DenseNet architecture is shown in Fig. 1a. The specific dimension of input data was $121\times145\times121$, in our applications. The output to be predicted was a single scalar representing chronological age (years). The architecture consisted of a regular $5\times5\times5$ convolutional layer followed by 4 dense blocks and 3 transition blocks in between them. The 4 dense blocks consisted of 3, 6, 12 and 8 dense layers, respectively (denoted above each block). Each dense layer had a $1\times1\times1$ bottleneck convolutional layer followed by a $3\times3\times3$ convolution layer. The dense layers were densely interconnected in a feed-forward manner within each block. The growth rate (k) was 48. The flattened output from the last global average pooling layer was then fully connected with 1,457 units and was connected to the output layer.

The neural network was implemented using Keras with TensorFlow[68] as the backend. Cross-validated experiments were conducted using fivefold validations (60% training dataset, 20% validation dataset and 20% test dataset). The MAE was used as the loss function. The model was optimized using the Adam optimizer with the parameters: $\beta1=0{:}9$ and $\beta2=0.99$ (ref. [69]). The He initialization strategy was used for weight initialization[70]. The training epoch was 150. The learning rate selected for the training set was 1×10^{-4} and decreased by a factor of 2 for every 10 epochs. If the validation error did not improve in seven epochs, the learning rate was updated. The hyperparameters were optimized based on the performance of the validation set in the hyperparameter tuning stage from an initial grid parameter search (batch size: (2, 4) and learning rate: (1×10^{-1}, 1×10^{-2}, 1×10^{-3}, 1×10^{-4}, 1×10^{-5})). The loss function, optimizer and learning rate scheduler and early stopping callback were fixed throughout the tuning stage. The total number of parameters were 70,183,073, of which 70,122,657 were trainable parameters. We used a mini-batch size of four. Training and testing were performed on a Tesla P100 GPU.

For a comparison with DenseNet, we also utilized 3D-ResNet101 (ref. [71]) and SFCN architectures[31]. We only compared the performance with other deep learning models, although non-deep learning models can offer the benefit of greater

Table 1 | Demographics of the Mayo dataset

Characteristic	Clinical diagnosis				
	Normal	MCI	AD	FTD	DLB
n	1,805	480	215	45	86
Total time points, *n* (%)					
1	973 (53.91)	190 (39.58)	80 (37.21)	19 (42.22)	44 (51.16)
2	503 (27.87)	130 (27.08)	72 (33.49)	10 (22.22)	15 (17.44)
3	243 (13.46)	86 (17.92)	31 (14.42)	8 (17.78)	10 (11.63)
4+	86 (4.76)	74 (15.42)	32 (14.88)	8 (17.78)	17 (19.76)
Age, years					
Median (IQR)	72 (62–79)	77 (70–83)	74 (64–79.75)	63 (55–70.25)	71 (66–77)
Minimum-maximum	30–97	26–98	49–92	31–76	45–90
Male sex, *n* (%)	952 (52.74)	319 (66.46)	117 (54.42)	26 (57.78)	74 (86.05)
Education, years, median (IQR)	15 (13–17)	14 (12–16)	16 (12–17.75)	16 (13–17.25)	15.5 (13–18)
CDR-SB, median (IQR)	0 (0)	0.5 (0.5–0.5)	1 (0.5–1)	1 (0.5–1)	1 (0.5–1)
MMSE, median (IQR)	29 (28–29)	27 (24–28)	21 (17–24)	24 (21–26)	23 (17–25.25)
STMS, median (IQR)	36 (34–37)	32 (29–34)	25 (19–29)	28 (25–31.5)	27 (22–30)

IQR, interquartile range.

interpretability because it has been well established that deep learning models perform far better than a non-deep learning approach and feature extraction may be problematic for non-deep learning models. For the ResNet training, we implemented an ADAM optimizer with an initial learning rate of 0.001. For the SFCN, we implemented a stochastic gradient descent optimizer with an initial learning rate of 0.01 and L2 weight decay parameter of 0.001. The source code is available at https://github.com/Neurology-AI-Program/Brain_age_prediction.git.

Occlusion sensitivity analysis. To facilitate interpretability, we generated brain maps of the relevant features used in the age prediction model using occlusion sensitivity analysis[72]. The analysis was conducted within the test dataset. To calculate the age-specific saliency map, data were separated into 7 sub-age groups based on their chronological age, from 30 to 100 with 10 year intervals. Within each group, the original images were occluded by $11\times11\times11$ voxel areas with zero values, along a $11\times11\times11$ grid (Fig. 1b). Since the front and rear 12 voxels along the anterior-posterior axes do not include the brain area, those were excluded from occlusion to reduce the computational load. Then, age inference on the occluded images was performed through our pretrained 3D-DenseNet model and performance was evaluated as $MAE_{occlusion}$. The delta MAE was obtained by calculating the difference between $MAE_{occlusion}$ and $MAE_{original}$ acquired through the original image; a delta MAE matrix ($11\times11\times11$) was obtained by iterating occlusion for every region ($n=1{,}331$). Then, the delta MAE matrix was reconstructed into the original image size ($121\times145\times121$) through cubic interpolation and zero padding for the excluded area in the occlusion; the average of the 5 folds was calculated. Normalization was performed by dividing the entire image by the maximum value; thus, the values of the final saliency map ranged from 0 to 1.

Dataset split experiment. To measure how the inclusion of multiple time points per participant affected brain age prediction, we tested five different data split options. The main result was derived from the strictest data split option: option 1 using only a single time point per participant. Four additional options were tested; option 2 (multiple time points per participant with overlap between training, validation and test datasets permitted); option 3 (multiple time points per participant with overlap between training and validation datasets permitted); option 4 (multiple time points for the training and validation datasets and a single time point for the test dataset; no overlap of participants among training, validation and test datasets were permitted); and option 5 (a single time point was used for the validation and test datasets; no overlap of participants among training, validation and test datasets). For these five options, the validation and test MAE from fivefold cross-validations were compared with option 1 (Supplementary Table 4).

Statistical analysis. Brain age prediction accuracy was assessed by MAE and Spearman's correlation between predicted age and chronological age. Defining x to be chronological age and y the predicted age, the brain age gap was calculated by $y-x$. The brain age gap is known to be correlated with chronological age, which results in an overestimation for younger individuals and an underestimation for older individuals[21,31] due to regression dilution[29]. Therefore, we used the linear bias correction method described in Smith et al.[30] for age bias correction for the brain age gap. We fitted a linear regression $y=ax+b$ to the test dataset. Then, the corrected brain age gap was calculated by $(y-b)/a-x$. The a and b coefficient derived from the cognitively unimpaired group was applied to other diagnostic groups in the same way for the bias correction. The corrected brain age gap of the disease groups was compared with the cognitively unimpaired participants by an one-way analysis of variance (ANOVA) with Holm-Šídák post hoc test. The Pearson's correlation coefficient was evaluated to test for an association between the FDG- and MRI-based brain age gap. These correlation coefficients were then compared between the cognitively unimpaired and disease groups using a z-test after Fisher's r-to-z transformation. The Pearson's correlation coefficient was utilized to test for an association of the corrected brain age gap and cognitive scores, and meta-ROI amyloid- and tau-PET SUVR. A pair-wise comparison of corrected brain age gap between females and males within clinical groups was evaluated using a two-sample Student's t-test. For the longitudinal analysis, the disease progression group seen in participants at the time points of the serial scans were defined as cognitively unimpaired to cognitively unimpaired, cognitively unimpaired to MCI/AD, MCI to MCI, MCI to AD, MCI to FTD, FTD to FTD, MCI to DLB and DLB to DLB, with the second category representing the most recent diagnostic group assignment. Then, the baseline (that is, earlier time point) brain age gap was compared within the same baseline groups (that is, cognitively unimpaired baseline: cognitively unimpaired to cognitively unimpaired versus cognitively unimpaired to MCI/AD; MCI baseline: MCI to MCI versus MCI to AD, MCI to FTD and MCI to DLB) using a one-way ANOVA with Holm-Šídák post hoc test. An annualized Δ brain age gap was calculated by a difference of brain age gaps between baseline and follow-up scan and dividing by the time difference in years. The annualized Δ brain age gap of each group was then compared to that of CU to CU group using an one-way ANOVA with Holm-Šídák post hoc test. A voxel-wise regression analysis was performed using the brain age gap as a regressor to investigate which brain regions' alterations were associated with brain age gap generation for each patient group. Each individual's chronological age was specified as a nuisance covariance. For cognitively unimpaired participants, the same analysis was performed using chronological age as a regressor. Statistical significance was corrected for multiple comparisons using a false discovery rate (FDR)[73] with a cluster size of at least 100 adjacent voxels. Three-dimensional models of the cortical surfaces were reconstructed with a standard recon-all command (FreeSurfer v 7.1.1)[74]. Surfaces were visualized using the SUMA software (https://afni.nimh.nih.gov/Suma). A similarity of beta map between the cognitively unimpaired and disease groups was evaluated using a voxel-wise Pearson's correlation analysis. These correlation coefficients were then compared between disease groups using a z-test after Fisher's r-to-z transformation. All analyses were performed with MATLAB v.9.4 (MathWorks) and Prism v. 9.1.2 (GraphPad Software).

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

Data availability

The Mayo dataset that supports the findings of this study is not publicly available. Anonymized data are available from the corresponding author upon reasonable request. The MRI and PET data from ADNI are available to researchers via the data access procedure described at http://adni.loni.usc.edu/data-samples/access-data/. Source data are provided with this paper.

Code availability

The code developed for this work is available at https://github.com/Neurology-AI-Program/Brain_age_prediction.git.

Received: 11 August 2021; Accepted: 29 March 2022;
Published online: 09 May 2022

References

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
2. Harman, D. Aging: overview. *Ann. N. Y. Acad. Sci.* **928**, 1–21 (2001).
3. Courchesne, E. et al. Normal brain development and aging: quantitative analysis at in vivo MR imaging in healthy volunteers. *Radiology* **216**, 672–682 (2000).
4. Good, C. D. et al. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* **14**, 21–36 (2001).
5. Sowell, E. R. et al. Mapping cortical change across the human life span. *Nat. Neurosci.* **6**, 309–315 (2003).
6. Lemaitre, H. et al. Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiol. Aging* **33**, 617.e1–617.e9 (2012).
7. Raz, N. & Rodrigue, K. M. Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neurosci. Biobehav. Rev.* **30**, 730–748 (2006).
8. Walhovd, K. B. et al. Effects of age on volumes of cortex, white matter and subcortical structures. *Neurobiol. Aging* **26**, 1261–1270 (2005).
9. Goyal, M. S. et al. Loss of brain aerobic glycolysis in normal human aging. *Cell Metab.* **26**, 353–360.e3 (2017).
10. Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z. & Raichle, M. E. Aerobic glycolysis in the human brain is associated with development and neotenous gene expression. *Cell Metab.* **19**, 49–57 (2014).
11. Zuendorf, G., Kerrouche, N., Herholz, K. & Baron, J.-C. Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional imaging data sets as applied to FDG-PET and normal aging. *Hum. Brain Mapp.* **18**, 13–21 (2003).
12. Knopman, D. S. et al. ^{18}F-fluorodeoxyglucose positron emission tomography, aging, and apolipoprotein E genotype in cognitively normal persons. *Neurobiol. Aging* **35**, 2096–2106 (2014).
13. De Leon, M. et al. Prediction of cognitive decline in normal elderly subjects with 2-[^{18}F]fluoro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET). *Proc. Natl Acad. Sci. USA* **98**, 10966–10971 (2001).
14. De Santi, S. et al. Age-related changes in brain: II. Positron emission tomography of frontal and temporal lobe glucose metabolism in normal subjects. *Psychiatr. Q.* **66**, 357–370 (1995).
15. Bonte, S. et al. Healthy brain ageing assessed with ^{18}F-FDG PET and age-dependent recovery factors after partial volume effect correction. *Eur. J. Nucl. Med. Mol. Imaging* **44**, 838–849 (2017).
16. Shen, X., Liu, H., Hu, Z., Hu, H. & Shi, P. The relationship between cerebral glucose metabolism and age: report of a large brain PET data set. *PLoS ONE* **7**, e51517 (2012).
17. Petit-Taboué, M., Landeau, B., Desson, J. F., Desgranges, B. & Baron, J. C. Effects of healthy aging on the regional cerebral metabolic rate of glucose assessed with statistical parametric mapping. *Neuroimage* **7**, 176–184 (1998).
18. Goyal, M. S. et al. Persistent metabolic youth in the aging female brain. *Proc. Natl Acad. Sci. USA* **116**, 3251–3255 (2019).
19. Cole, J. H. & Franke, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci.* **40**, 681–690 (2017).
20. Cole, J. H. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. *Neurobiol. Aging* **92**, 34–42 (2020).
21. Bashyam, V. M. et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* **143**, 2312–2324 (2020).
22. Abrol, A. et al. Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nat. Commun.* **12**, 353 (2021).
23. Cole, J. H. et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage* **163**, 115–124 (2017).
24. Levakov, G., Rosenthal, G., Shelef, I., Raviv, T. R. & Avidan, G. From a deep learning model back to the brain—identifying regional predictors and their relation to aging. *Hum. Brain Mapp.* **41**, 3235–3252 (2020).
25. Jonsson, B. A. et al. Brain age prediction using deep learning uncovers associated sequence variants. *Nat. Commun.* **10**, 5409 (2019).
26. Cole, J. H. et al. Brain age predicts mortality. *Mol. Psychiatry* **23**, 1385–1392 (2018).
27. Gaser, C. et al. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. *PLoS ONE* **8**, e67346 (2013).
28. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional Networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (eds Rehg, J. et al.) 4700–4708. (Institute of Electrical and Electronics Engineers, 2017).
29. MacMahon, S. et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* **335**, 765–774 (1990).
30. Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *Neuroimage* **200**, 528–539 (2019).
31. Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. & Smith, S. M. Accurate brain age prediction with lightweight deep neural networks. *Med. Image Anal.* **68**, 101871 (2021).
32. Morris, J. C. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int. Psychogeriatr.* **9**, 173–176 (1997).
33. Kokmen, E., Smith, G. E., Petersen, R. C., Tangalos, E. & Ivnik, R. C. The short test of mental status: correlations with standardized psychometric testing. *Arch. Neurol.* **48**, 725–728 (1991).
34. Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
35. Chételat, G. et al. Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease. *Brain* **131**, 60–71 (2008).
36. Salat, D. H. et al. Thinning of the cerebral cortex in aging. *Cereb. Cortex* **14**, 721–730 (2004).
37. Buckner, R. L. et al. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* **25**, 7709–7717 (2005).
38. Curiati, P. K. et al. Age-related metabolic profiles in cognitively healthy elders: results from a voxel-based [^{18}F]fluorodeoxyglucose–positron-emission tomography study with partial volume effects correction. *AJNR Am. J. Neuroradiol.* **32**, 560–565 (2011).
39. Long, X. et al. Healthy aging: an automatic analysis of global and regional morphological alterations of human brain. *Acad. Radiol.* **19**, 785–793 (2012).
40. Jack, C. R. Jr. et al. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* **51**, 993–999 (1998).
41. Davis, P. C., Mirra, S. S. & Alazraki, N. The brain in older persons with and without dementia: findings on MR, PET, and SPECT images. *AJR Am. J. Roentgenol.* **162**, 1267–1278 (1994).
42. Habes, M. et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain* **139**, 1164–1179 (2016).
43. Ossenkoppele, R. et al. Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease. *Neurology* **92**, e601–e612 (2019).
44. Franke, K. & Gaser, C. Longitudinal changes in individual BrainAGE in healthy aging, mild cognitive impairment, and Alzheimer's disease. *GeroPsych (Bern)* **25**, 235–245 (2012).
45. Shivamurthy, V. K., Tahari, A. K., Marcus, C. & Subramaniam, R. M. Brain FDG PET and the diagnosis of dementia. *AJR Am. J. Roentgenol.* **204**, W76–W85 (2015).
46. Brown, R. K., Bohnen, N. I., Wong, K. K., Minoshima, S. & Frey, K. A. Brain PET in suspected dementia: patterns of altered FDG metabolism. *Radiographics* **34**, 684–701 (2014).
47. Kanda, T. et al. Comparison of grey matter and metabolic reductions in frontotemporal dementia using FDG-PET and voxel-based morphometric MR studies. *Eur. J. Nucl. Med. Mol. Imaging* **35**, 2227–2234 (2008).
48. Castelnovo, V. et al. Heterogeneous brain FDG-PET metabolic patterns in patients with C9orf72 mutation. *Neurol. Sci.* **40**, 515–521 (2019).
49. McKeith, I. G. et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. *Neurology* **89**, 88–100 (2017).
50. Graff-Radford, J. et al. ^{18}F-fluorodeoxyglucose positron emission tomography in dementia with Lewy bodies. *Brain Commun.* **2**, fcaa040 (2020).
51. Hayflick, L. Biological aging is no longer an unsolved problem. *Ann. N. Y. Acad. Sci.* **1100**, 1–13 (2007).
52. Berg, L. Does Alzheimer's disease represent an exaggeration of normal aging? *Arch. Neurol.* **42**, 737–739 (1985).
53. Toepper, M. Dissociating normal aging from Alzheimer's disease: a view from cognitive neuroscience. *J. Alzheimers Dis.* **57**, 331–352 (2017).
54. Rieke, J., Eitel, F., Weygandt, M., Haynes, J.-D. & Ritter, K. Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease. In *Proc. Understanding and Interpreting Machine Learning in Medical Image Computing Applications* (eds Stoyanov, D. et al.) 24–31 (Springer, 2018).
55. Jones, D. T. et al. Tau, amyloid, and cascading network failure across the Alzheimer's disease spectrum. *Cortex* **97**, 143–159 (2017).
56. Roberts, R. O. et al. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology* **30**, 58–69 (2008).

57. Albert, M. S. et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 270–279 (2011).
58. McKhann, G. M. et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 263–269 (2011).
59. Petersen, R. C. Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* **256**, 183–194 (2004).
60. Neary, D. et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* **51**, 1546–1554 (1998).
61. Klunk, W. E. et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann. Neurol.* **55**, 306–319 (2004).
62. Xia, C.-F. et al. [^{18}F]T807, a novel tau positron emission tomography imaging agent for Alzheimer's disease. *Alzheimers Dement.* **9**, 666–676 (2013).
63. Schwarz, C. G. et al. A comparison of partial volume correction techniques for measuring change in serial amyloid PET SUVR. *J. Alzheimers Dis.* **67**, 181–195 (2019).
64. Schwarz, C. et al. The mayo clinic adult lifespan template (MCALT): better quantification across the lifespan. In *Proc. Alzheimer's Association International Conference* 13: P792. (The Alzheimer's Association, 2017).
65. Ashburner, J. & Friston, K. J. Unified segmentation. *Neuroimage* **26**, 839–851 (2005).
66. Jack, C. R. Jr. et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement.* **13**, 205–216 (2017).
67. Shinohara, R. T. et al. Statistical normalization techniques for magnetic resonance imaging. *Neuroimage Clin.* **6**, 9–19 (2014).
68. Abadi, M. et al. Tensorflow: A system for large-scale machine learning. In *Proc. 12th (USENIX) Symposium on Operating Systems Design and Implementation* (eds Keeton, K. & Roscoe, T.) 265–283. (USENIX Association, 2016).
69. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1412.6980 (2014).
70. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification In *Proc. IEEE International Conference on Computer Vision* (eds Ikeuchi, K. et al.) 1026–1034. (Institute of Electrical and Electronics Engineers, 2015).
71. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (eds Agapito, L. et al.) 770–778. (Institute of Electrical and Electronics Engineers, 2016).
72. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *Proc. European Conference on Computer Vision* (eds Fleet, D. et al.) 818–833 (Springer, 2014).
73. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).
74. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

Acknowledgements

We acknowledge the support of the NVIDIA Corporation with the donation of the Tesla P100 GPU used for this research. This work was funded in part by National Institutes of Health grant nos. P30 AG62677 (D.J.), R01 AG011378 (C.J.), R01 AG041851 (C.J.), P50 AG016574 (R.P.), U01 AG06786 (R.P.) and RO1 AG073282 (V.L.), and by the Elsie and Marvin Dekelboum Family Foundation, Edson Family Foundation, Liston Family Foundation, GHR Foundation, Foundation Dr. Corinne Schuler (Geneva, Switzerland), Race Against Dementia and the Mayo Foundation.

Author contributions

J.L., H.K.M. and D.T.J. conceptualized the study. J.L. and L.R.B. were responsible for the software. M.L.S. was responsible for preprocessing the data. J.L., H.K.M., M.L.S., E.S.L., H.B., J.G., J.L.G. and C.G.S. were responsible for the study methodology. J.L., B.J.B., H.K.M., V.J.L. and D.T.J. wrote the original manuscript. All authors revised the draft. K.K., D.S.K., B.F.B., V.J.L., R.C.P., C.R.J. and D.T.J. supervised the study. All authors gave final approval for this version of the article.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-022-00219-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-022-00219-7.

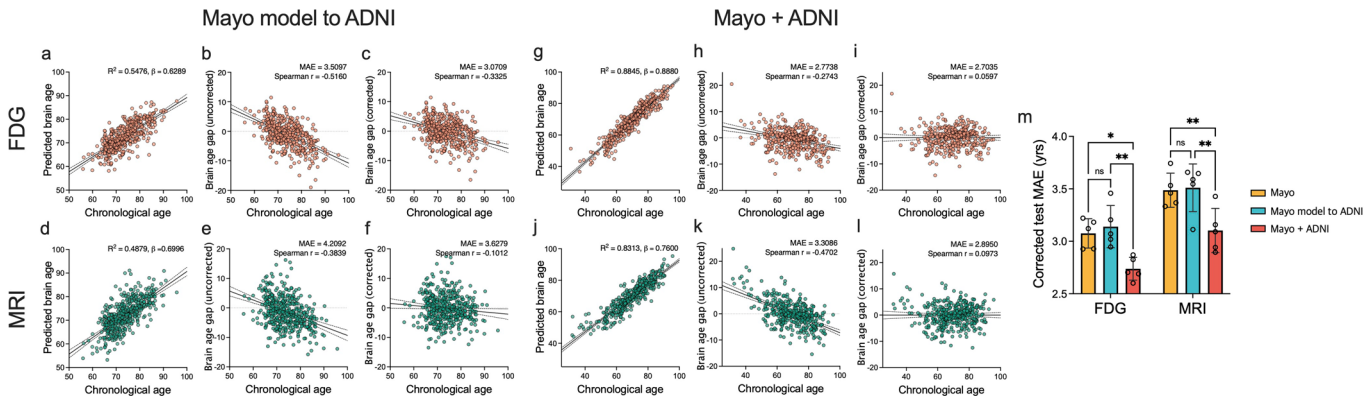
Correspondence and requests for materials should be addressed to David T. Jones.

Peer review information *Nature Aging* thanks Ahmad Hariri, Brian Gordon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

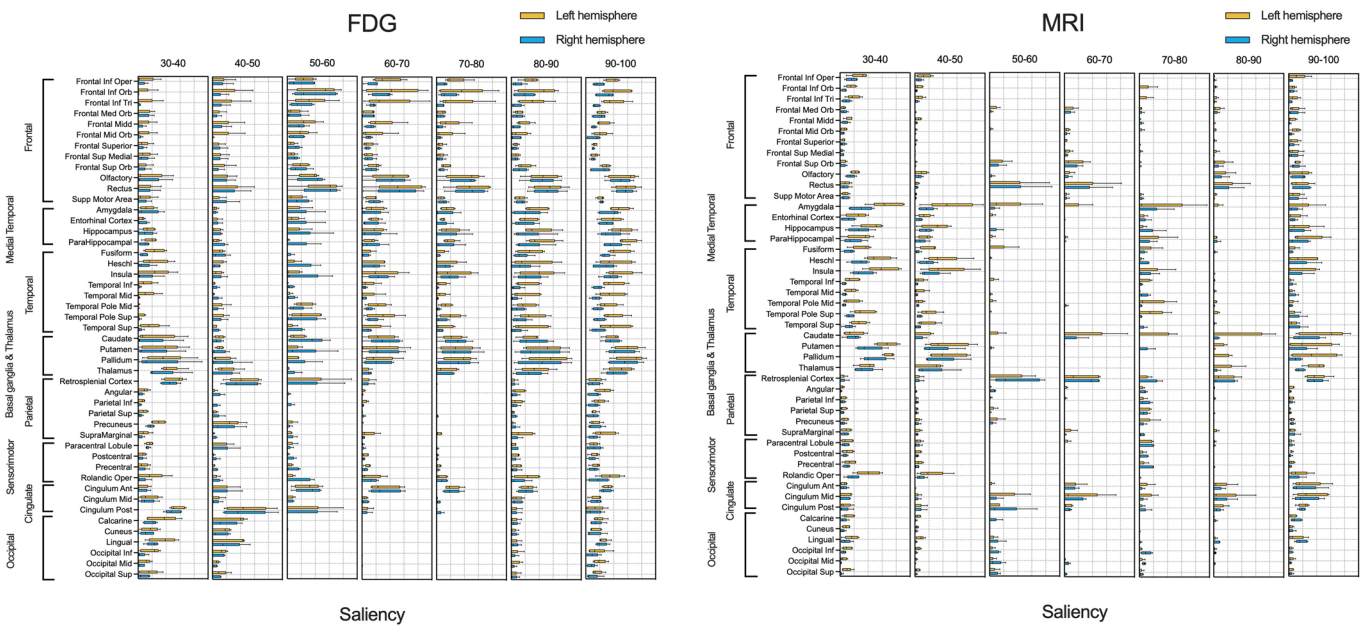
Reprints and permissions information is available at www.nature.com/reprints.

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

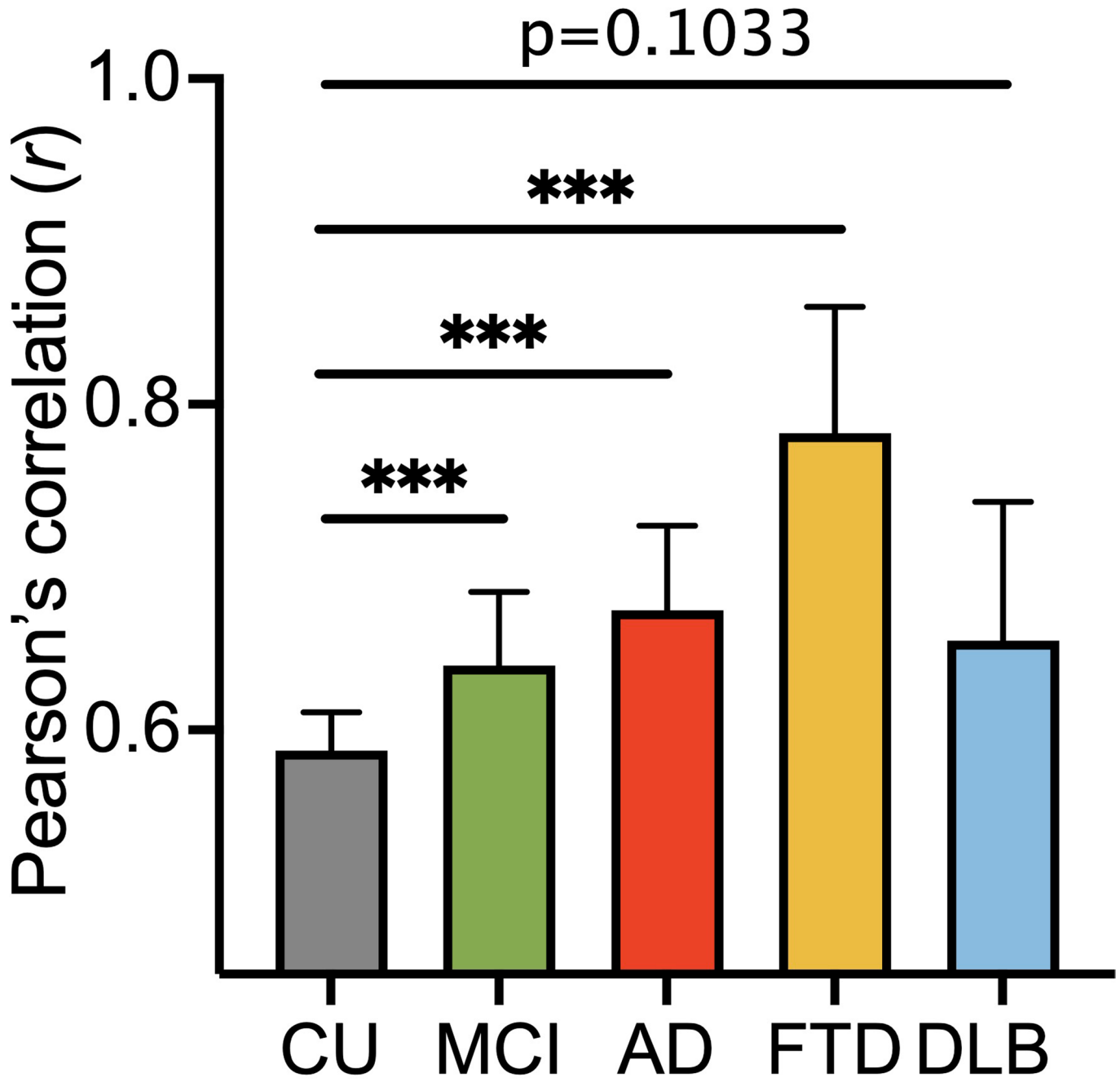
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022



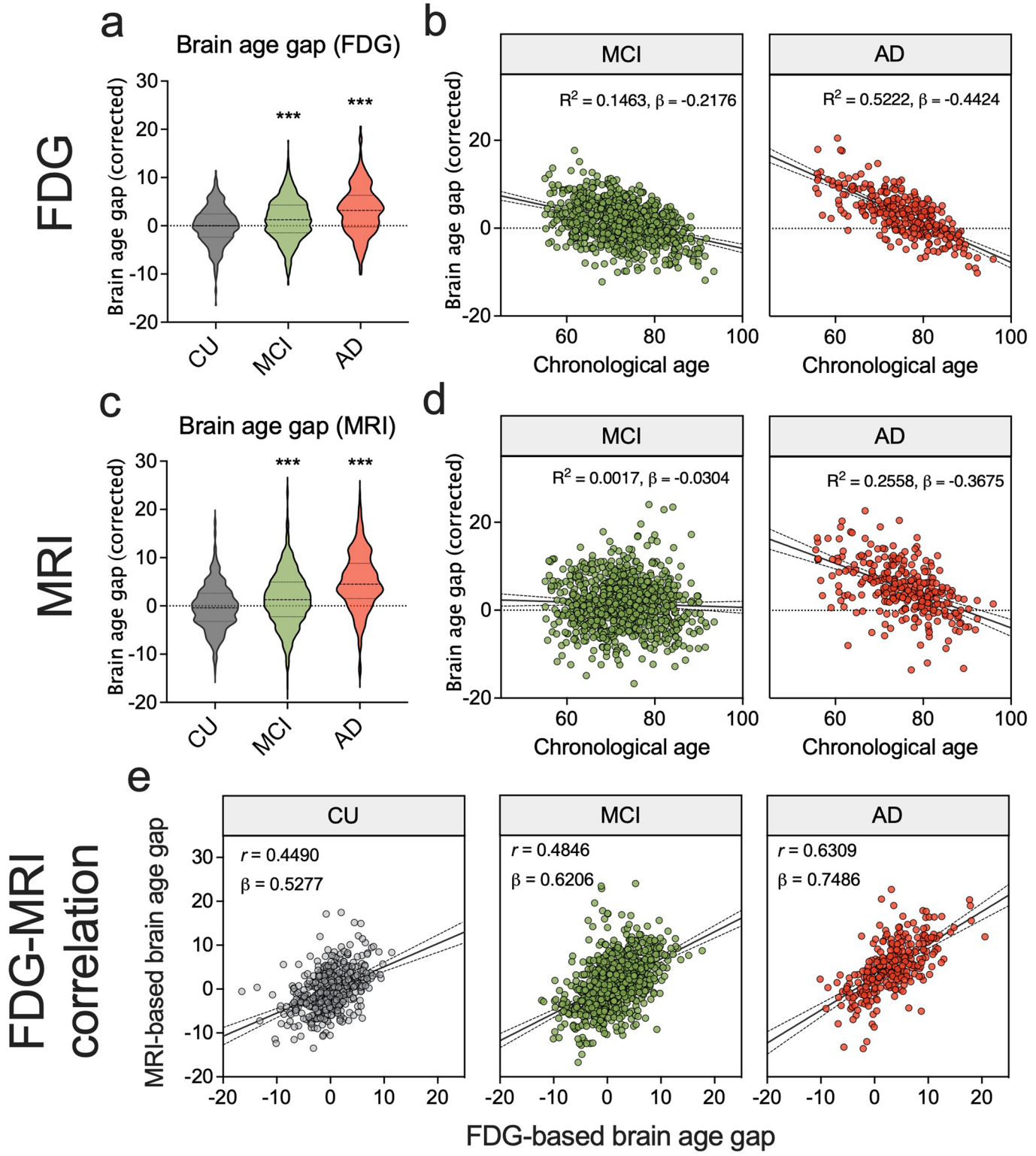
Extended Data Fig. 1 | Brain age predictions on the ADNI dataset. (**a**-**f**) 3D Densenet model trained on the Mayo dataset was applied to the ADNI data. (a-c) FDG based brain age prediction. (**d**-**f**) MRI based brain age prediction. (g-l) Prediction performance of 3D Densenet model trained on the Mayo and ADNI dataset together. (g-i) FDG based brain age prediction. (**j**-**l**) MRI based brain age prediction. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. **m**, The corrected MAE evaluated on the test data (n = 5) was compared between the datasets using a two−sided two-sample Student's t-test. The data is shown as mean ± SD. * $p < 0.05$, ** $p < 0.005$.



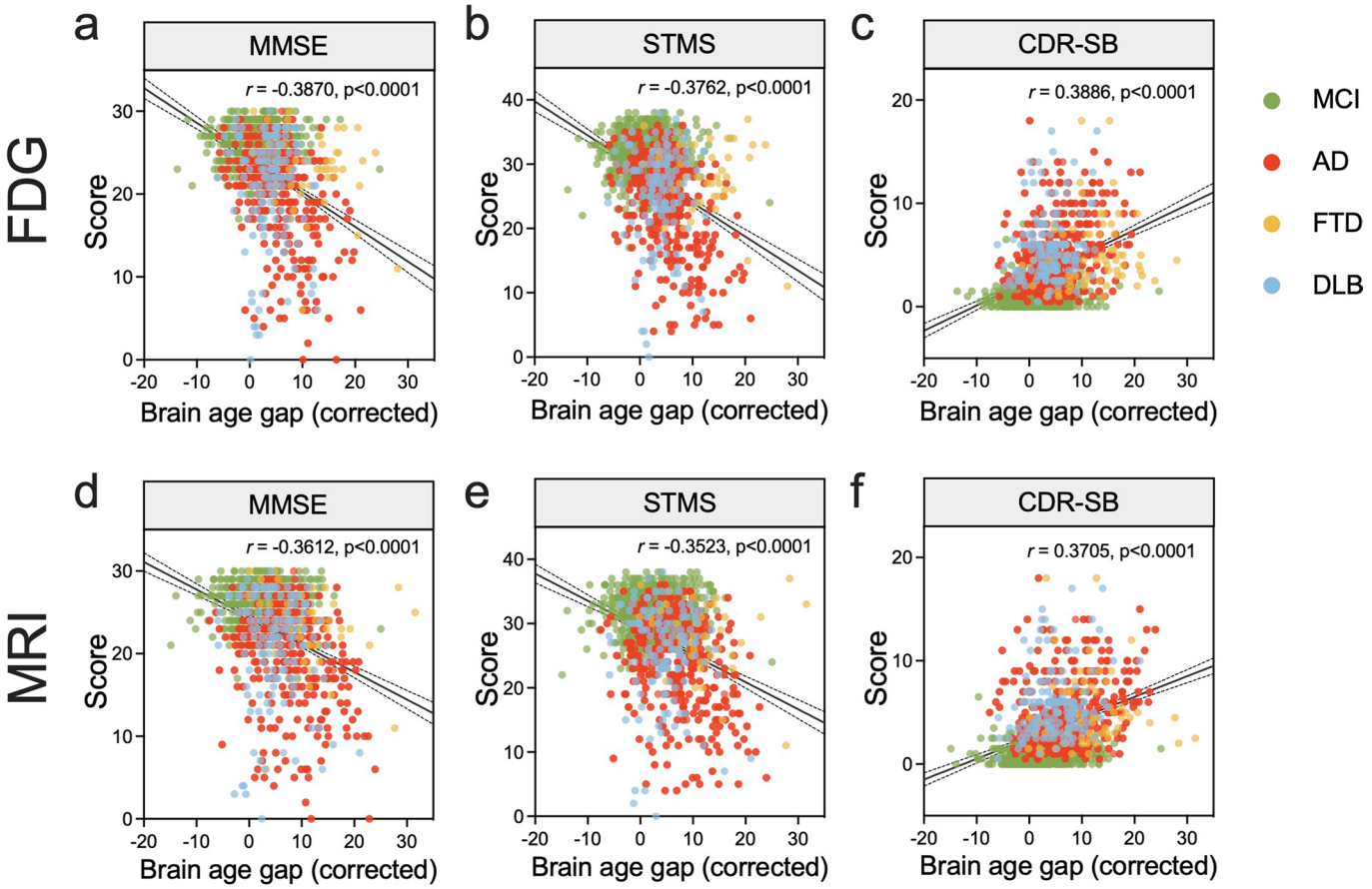
Extended Data Fig. 2 | Regional mean saliency. After calculating the saliency map from occlusion analysis, mean saliency value was calculated for each ROI. Box plots represent the minimum and maximum values (whiskers), the first and third quartile (box boundaries), and the median (internal line) for the 5-fold cross-validations (n = 5). Yellow-colored boxes indicate the left hemisphere and blue-colored boxes indicate the right hemisphere.



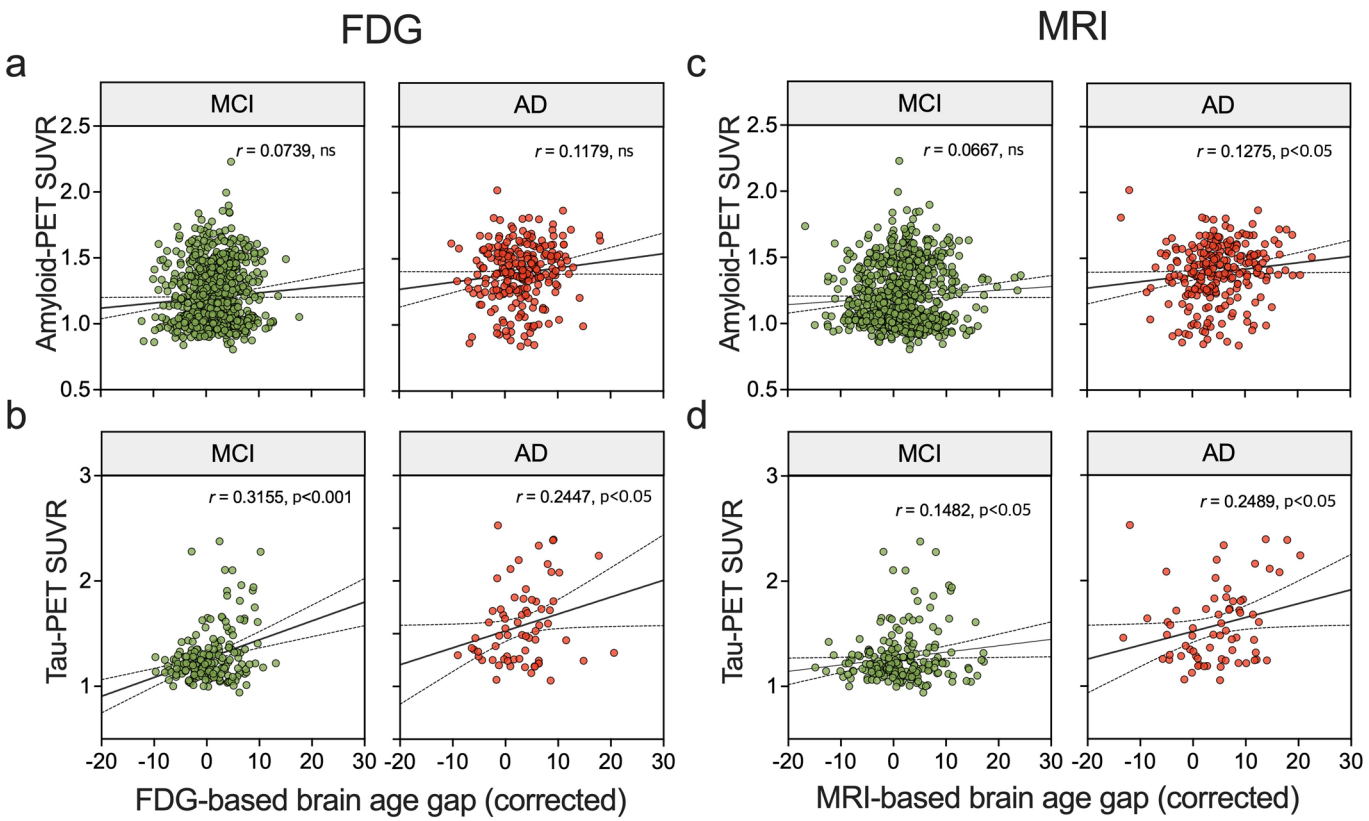
Extended Data Fig. 3 | Comparison of correlation between FDG- and MRI-based brain age gap. Error bars indicate 95% confidence intervals of Pearson's correlation coefficient. A statistical comparison was performed with the CU group. Pearson's r (95% confidence interval) = 0.5873 (0.5628 to 0.6108), 0.6396 (0.5945 to 0.6847), 0.6735 (0.6138 to 0.7255), 0.7824 (0.6697 to 0.8598), and 0.6548 (0.5489 to 0.7400), for CU, MCI, AD, FTD, and DLB, respectively. Exact p values: CU versus MCI, $p = 1.5 \times 10^{-9}$; CU versus AD, $p < 1 \times 10^{-15}$; CU versus FTD, $p < 1 \times 10^{-15}$; CU versus DLB, $p < 1 \times 10^{-15}$, *** $p < 0.001$. *** $p < 0.001$, two-sided z test after Fisher's r to z transformation.



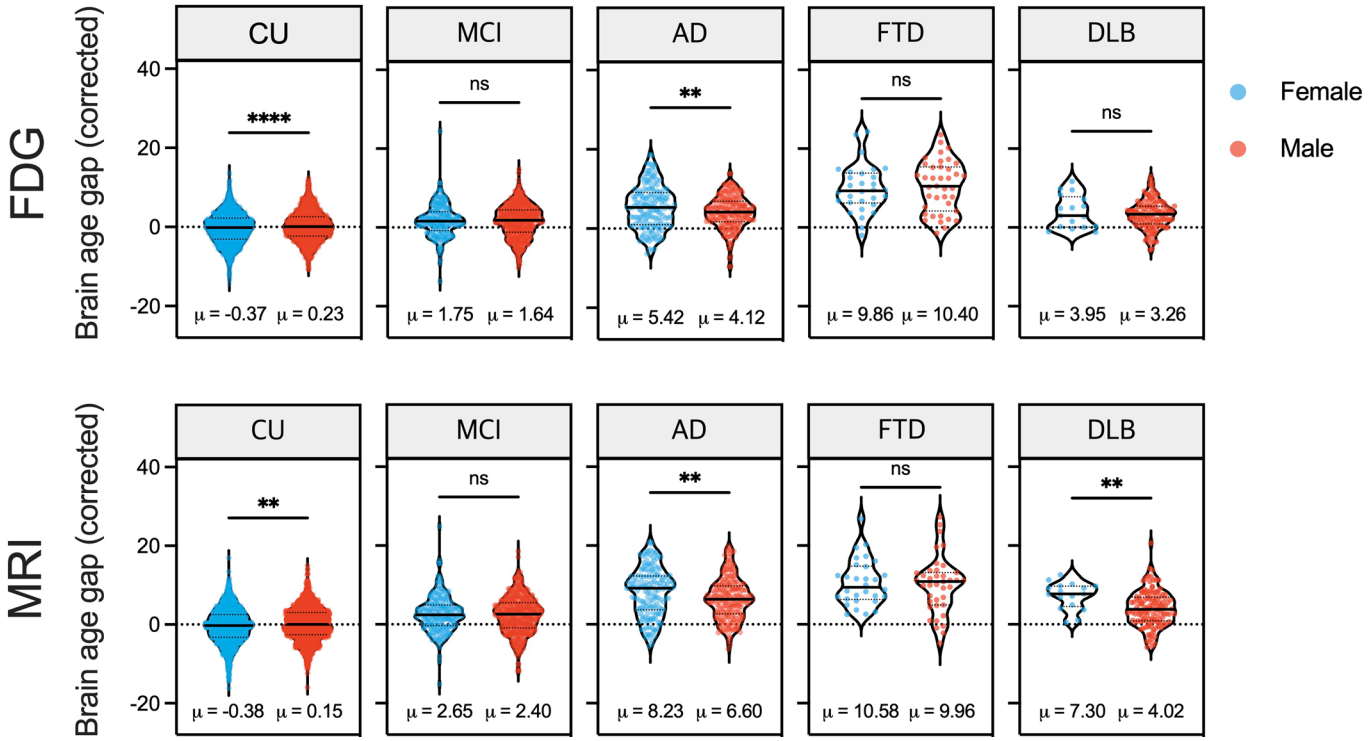
Extended Data Fig. 4 | Regression plots of a corrected brain age gap as a function of chronological age for clinical diagnostic groups in ADNI cohort. **a**, Violin plots of corrected brain age gap for each diagnostic group. The corrected brain age gap of disease groups was compared with CU using one-way ANOVA with Holm-Sidak's multiple comparisons test. *** $p < 0.001$. **b**, FDG-based brain age gap estimation for MCI and AD, respectively. **c**, Violin plots of corrected brain age gap for each clinical diagnosis group. The corrected brain age gap of disease groups was compared with CU using one-way ANOVA with Holm-Sidak's multiple comparisons test. *** $p < 0.001$. **d**, MRI-based brain age gap estimation for MCI and AD, respectively. **e**, Relationship between FDG- and MRI-based brain age gap. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. r indicates Pearson's correlation coefficient.



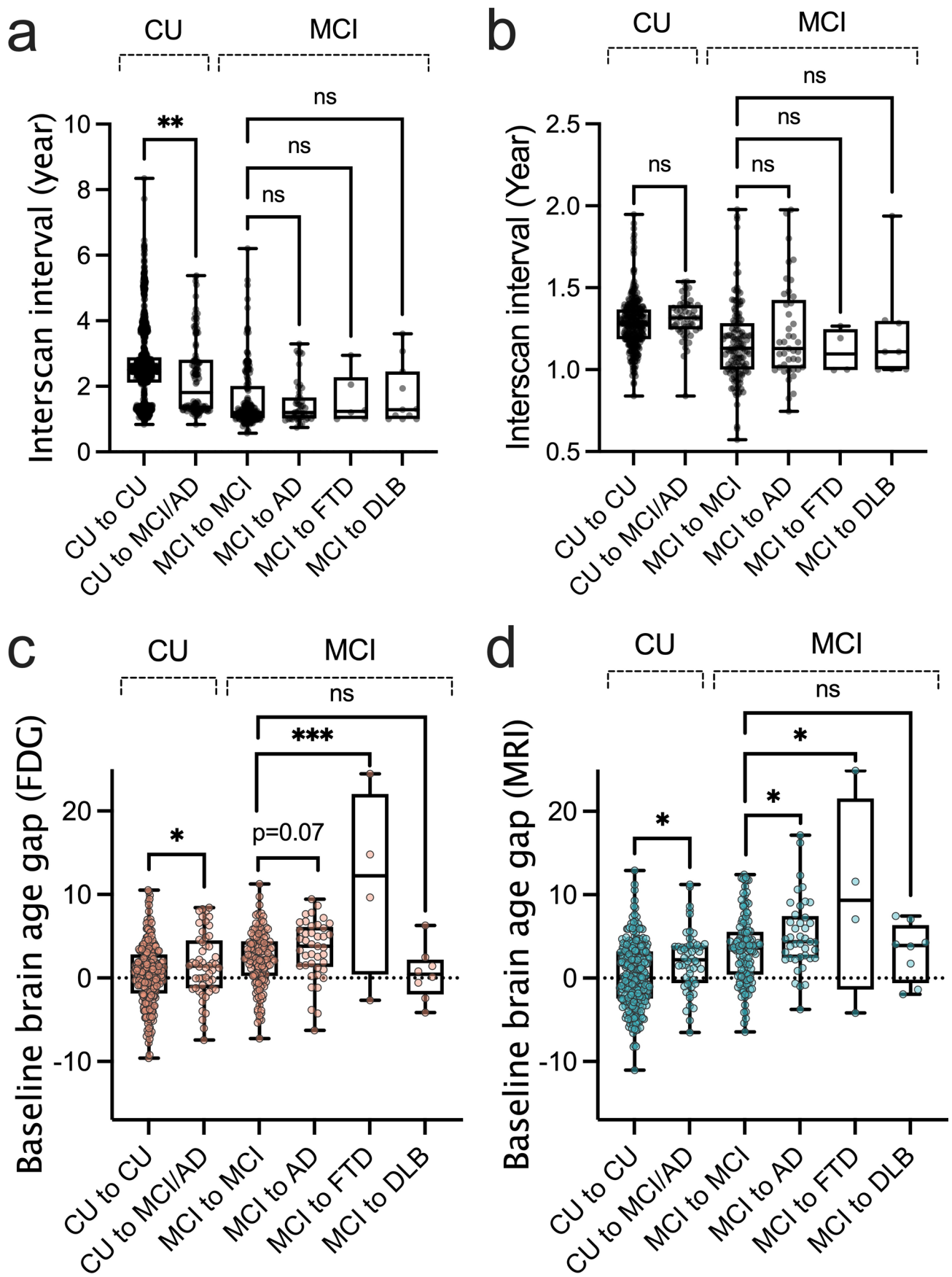
Extended Data Fig. 5 | Association of a brain age gap with cognitive scores. (**a**-**c**) Scatter plots of FDG model-based brain age gap with Mini-Mental State Examinations (MMSE), Short Test of Mental Status (STMS) and Clinical Dementia Rating Sum of boxes (CDR-SB), respectively. (**d**-**f**) Scatter plots of MRI model-based brain age gap with MMSE, STMS and CDR-SB, respectively. r, Pearson correlation coefficient; p, correlation test p value.



Extended Data Fig. 6 | Association of brain age gap with meta-ROI Amyloid- and Tau PET SUVr in ADNI cohort. a, Scatter plots between FDG-based brain age gap with meta-ROI amyloid PET SUVr for MCI and AD, respectively. **b**, Scatter plots between FDG-based brain age gap with meta-ROI tau PET SUVr. **c**, Scatter plots between MRI-based brain age gap with meta-ROI PiB PET SUVr. **d**, Scatter plots between MRI-based brain age gap with meta-ROI Tau PET SUVr. The black solid line and dotted lines in each figure represent a regression line and its 95% confidence bands, respectively. r, Pearson correlation coefficient; p, correlation test p value.

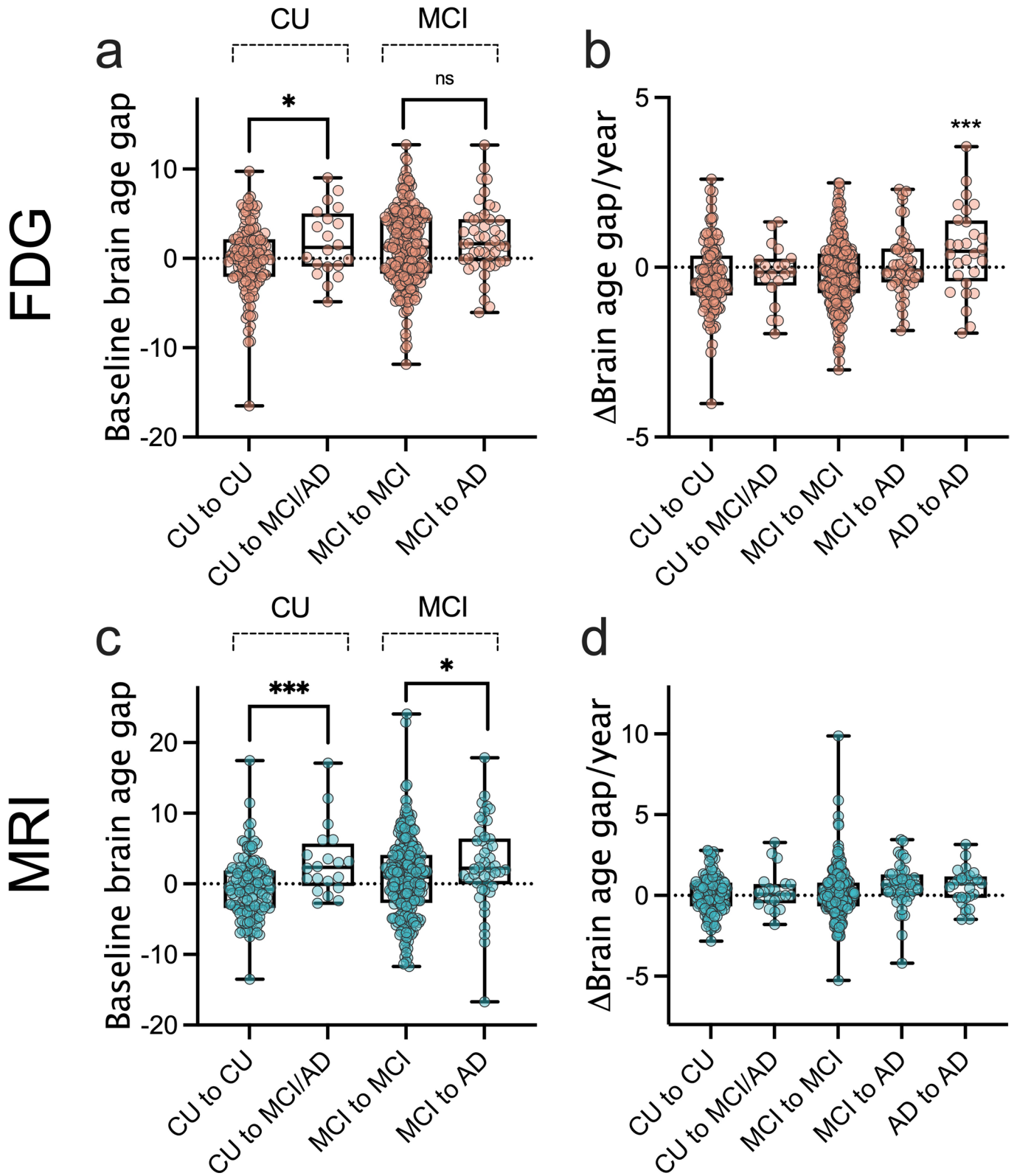


Extended Data Fig. 7 | Association of sex to the age gap estimation. The blue-colored dot indicates female and red indicates male individuals. Comparisons were calculated by two-sided Student's t-test. Exact p values: for FDG, CU, $t(2877) = 4.088$, $p = 4.5 \times 10^{-5}$; MCI, $t(664) = 0.3193$, $p = 0.7496$; AD, $t(370) = 2.625$, $p = 0.009$; FTD, $t(67) = 0.3496$, $p = 0.7277$; DLB, $t(139) = 0.7241$, $p = 0.4702$; for MRI, CU, $t(2877) = 3.290$, $p = 0.001$; MCI, $t(664) = 0.6509$, $p = 0.5154$; AD, $t(370) = 2.809$, $p = 0.0052$; FTD, $t(67) = 0.3811$, $p = 0.7043$; DLB, $t(139) = 2.886$, $p = 0.0045$; * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$.

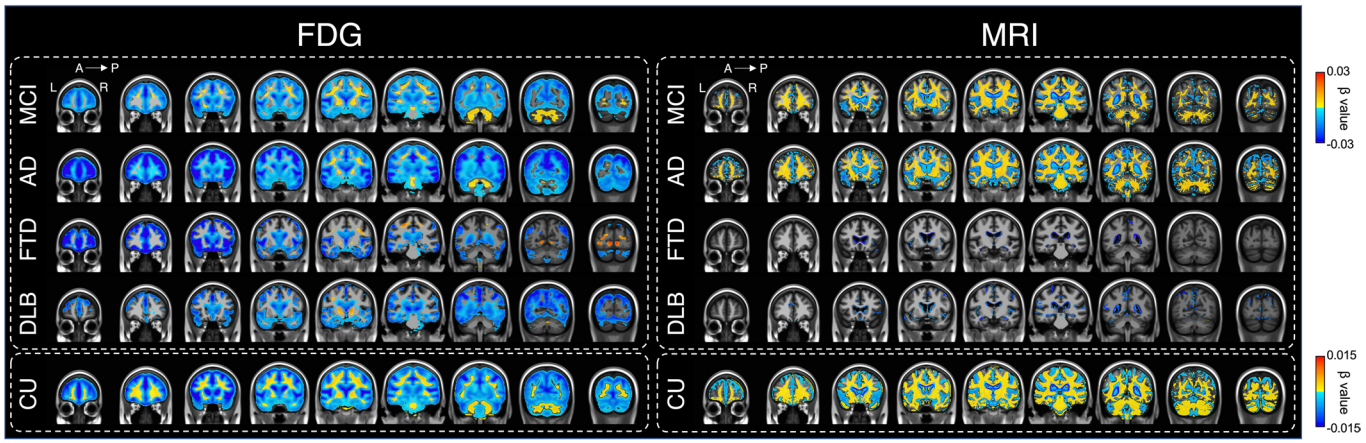


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Interscan interval bias test. a, Interscan interval for total subjects. A statistical test was performed within the same baseline groups (one-way ANOVA with Holm-Sidak post hoc test). Exact p values: CU to CU versus CU to MCI/AD, $p = 0.004$; MCI to MCI versus MCI to AD, $p = 0.79$; MCI to MCI versus MCI to FTD, $p = 0.95$; MCI to MCI versus MCI to DLB, $p = 0.95$. $n = 1054$, 104, 169, 49, 6, and 11 for CU to CU, CU to MCI/AD, MCI to MCI, MCI to AD, MCI to FTD, and MCI to DLB group, respectively. **b**, Interscan interval after excluding participants with interscan interval of >2 years (one-way ANOVA with Holm-Sidak post hoc test). Exact p values: CU to CU versus CU to MCI/AD, $p = 0.77$; MCI to MCI versus MCI to AD, $p = 0.29$; MCI to MCI versus MCI to FTD, $p = 0.77$; MCI to MCI versus MCI to DLB, $p = 0.77$. $n = 258$, 52, 127, 41, 4, and 8 for CU to CU, CU to MCI/AD, MCI to MCI, MCI to AD, MCI to FTD, and MCI to DLB group, respectively. (**c**,**d**) Baseline brain age gap comparison between groups after excluding participants with interscan interval of >2 years for FDG and MRI, respectively. The comparison was performed within the same baseline groups (one-way ANOVA with Holm-Sidak post hoc test). For c panel, Exact p values: CU to CU versus CU to MCI/AD, $p = 0.001$; MCI to MCI versus MCI to AD, $p = 0.07$; MCI to MCI versus MCI to FTD, $p = 3.2 \times 10^{-5}$; MCI to MCI versus MCI to DLB, $p = 0.38$. For d panel, Exact p values: CU to CU versus CU to MCI/AD, $p = 9.1 \times 10^{-4}$; MCI to MCI versus MCI to AD, $p = 0.02$; MCI to MCI versus MCI to FTD, $p = 0.12$; MCI to MCI versus MCI to DLB, $p = 0.94$. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$. $n = 258$, 52, 127, 41, 4, and 8 for CU to CU, CU to MCI/AD, MCI to MCI, MCI to AD, MCI to FTD, and MCI to DLB group, respectively. Box plots represent the minimum and maximum values (whiskers), the first and third quartile (box boundaries), and the median (internal line).



Extended Data Fig. 9 | Longitudinal nature of the brain age gap in ADNI cohort. a, Baseline brain age gap comparison between groups for FDG model. A statistical test was performed within the same baseline groups using a two-sided two-sample Student's t-test. **b**, For FDG model, the annual Δ brain age gap of each group was compared with the CU to CU group using one-way ANOVA with Holm-Sidak post hoc. **c**, Baseline brain age gap comparison between groups for MRI model. A statistical test was performed within the same baseline groups using a two-sided two-sample Student's t-test. **d**, For MRI model, the annual Δ brain age gap of each group was compared with the CU to CU group using one-way ANOVA with Holm-Sidak post hoc test). * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$. n = 124, 20, 237, 46, and 28 for CU to CU, CU to MCI/AD, MCI to MCI, MCI to AD, and AD to AD group, respectively. Box plots represent the minimum and maximum values (whiskers), the first and third quartile (box boundaries), and the median (internal line).



Extended Data Fig. 10 | Voxel-wise linear regression analysis of brain age gap shown on coronal slices. Clinical diagnosis group (MCI, AD, FTD and DLB)-specific results from voxel-wise whole-brain linear regression examining the brain age gap-related change (FDR corrected, $q < 0.01$). The chronological age was specified as nuisance covariance. For CU (bottom row), voxel-wise linear regression analysis was performed using the chronological age as a regressor to show the age-related change. A left panel shows the results for the FDG-based model and a right panel shows the results for the MRI-based model.

nature portfolio

Corresponding author(s): David T. Jones

Last updated by author(s): Mar 14, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Matlab (9.4.0), AFNI (20.0.18), SUMA (compile date: Feb. 26. 2020), Freesurfer (7.1.1), SPM12 and Prism (9.1.2) were used for data analysis and statistical test. The custom code used for model training in this article is made available at https://github.com/Neurology-AI-Program/Brain_age_prediction.git. The following packages are used: Tensorflow (1.9.0) keras (2.2.0) Conda (4.8.2) cudatoolkit (9.1.85) cudnn (7.0.5) numpy (1.19.5) python (3.6.10) scikit-learn (0.22.2) scipy (1.4.1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Mayo dataset that support the findings of this study are not publicly available. Anonymized data may be available from the corresponding author upon reasonable request and with permission. The MRI and PET data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) are available to researchers via data access procedure described at http://adni.loni.usc.edu/data-samples/access-data/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study did not require sample size estimation by power analysis. This study included the Mayo and ADNI cohorts. For the Mayo dataset, all available participants (n=2,349 and number of scans = 4,127) who had both 3T MRI and FDG PET from the Mayo Clinic Study of Aging (MCSA) or the Alzheimer's Disease Research Center (ADRC) study was included. For the ADNI dataset, all visits (n = 1,150 and number of scans = 1,622) with both 3T accelerated T1 and FDG PET scans was included.
Data exclusions	No data was excluded.
Replication	A cross-validation procedure (n=5) with train/validation/test partitions was used for all undertaken model training in this study. ADNI dataset was used to ensure reliability of the model's performance trained and tested on Mayo dataset.
Randomization	As this is a retrospective observational study design, randomization was not performed and is not applicable to our study.
Blinding	As this is a retrospective observational study design, blinding was not performed and is not applicable to our study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The Mayo cohort included 2,349 participants. The mean age was 70.83 years±11.38 ranging from 20 to 98. There were 1,037 females and 1,312 males. Table 1 shows characteristics of participants from the Mayo dataset. The ADNI cohort included 1,150 participants. The mean age was 74.05 years±7.55 ranging from 55 to 96. There were 540 females and 610 males. Supplementary table 1 shows characteristics of participants from the ADNI cohort.

Recruitment

All individuals in the Mayo cohort were enrolled in one of the two longitudinal cohort studies at Mayo Clinic: (i)The Mayo Clinic Study of Aging (MCSA), which is a longitudinal population-based study of cognitive aging among a stratified random sample of a geographically defined population (Olmsted County, Minnesota, USA); (ii) The Alzheimer's Disease Research Center (ADRC), which is a longitudinal research study of individuals recruited from the Mayo Clinic neurology practice.

For ADNI cohort, recruitment details may be found at http://adni.loni.usc.edu/about/.

Ethics oversight

All participants or designees provided written consent with the approval of Mayo Clinic and Olmsted Medical Center Institutional Review Boards.

The ADNI study procedures were approved by the institutional review boards of all participating centers as detailed in this document - https://adni.loni.usc.edu/wp-content/ uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Written, informed consent was obtained from all subjects participating in the study according to the Declaration of Helsinki. ADNI obtained all IRB approvals and met all ethical standards in the collection of data. The following are the ethics committees and IRB boards that provided approval. The Ethics committees/institutional review boards that approved the ADNI study are: Albany Medical Center Committee on Research Involving Human Subjects Institutional Review Board, Boston University Medical Campus and Boston Medical Center Institutional Review Board, Butler Hospital Institutional Review Board, Cleveland Clinic Institutional Review Board, Columbia University Medical Center Institutional Review Board, Duke University Health System Institutional Review Board, Emory Institutional Review Board, Georgetown University Institutional Review Board, Health Sciences Institutional Review Board, Houston Methodist Institutional Review Board, Howard University Office of Regulatory Research Compliance, Icahn School of Medicine at Mount Sinai Program for the Protection of Human Subjects, Indiana University Institutional Review Board, Institutional Review Board of Baylor College of Medicine, Jewish General Hospital Research Ethics Board, Johns Hopkins Medicine Institutional Review Board, Lifespan - Rhode Island Hospital Institutional Review Board, Mayo Clinic Institutional Review Board, Mount Sinai Medical Center Institutional Review Board, Nathan Kline Institute for Psychiatric Research & Rockland Psychiatric Center Institutional Review Board, New York University Langone Medical Center School of Medicine Institutional Review Board, Northwestern University Institutional Review Board, Oregon Health and Science University Institutional Review Board, Partners Human Research Committee Research Ethics, Board Sunnybrook Health Sciences Centre, Roper St. Francis Healthcare Institutional Review Board, Rush University Medical Center Institutional Review Board, St. Joseph's Phoenix Institutional Review Board, Stanford Institutional Review Board, The Ohio State University Institutional Review Board, University Hospitals Cleveland Medical Center Institutional Review Board, University of Alabama Office of the IRB, University of British Columbia Research Ethics Board, University of California Davis Institutional Review Board Administration, University of California Los Angeles Office of the Human Research Protection Program, University of California San Diego Human Research Protections Program, University of California San Francisco Human Research Protection Program, University of Iowa Institutional Review Board, University of Kansas Medical Center Human Subjects Committee, University of Kentucky Medical Institutional Review Board, University of Michigan Medical School Institutional Review Board, University of Pennsylvania Institutional Review Board, University of Pittsburgh Institutional Review Board, University of Rochester Research Subjects Review Board, University of South Florida Institutional Review Board, University of Southern, California Institutional Review Board, UT Southwestern Institution Review Board, VA Long Beach Healthcare System Institutional Review Board, Vanderbilt University Medical Center Institutional Review Board, Wake Forest School of Medicine Institutional Review Board, Washington University School of Medicine Institutional Review Board, Western Institutional Review Board, Western University Health Sciences Research Ethics Board, and Yale University Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type: Structural MRI (T1); no task was performed.

Design specifications: Only 3D high spatial resolution MRI were performed. No fMRI data was acquired.

Behavioral performance measures: Behavioral performance measures was not used in this study.

Acquisition

Imaging type(s): Structural MRI

Field strength: 3T

Sequence & imaging parameters: T1-weighted MRI scans were acquired using 3T scanners manufactured by General Electric (GE) and Siemens using a 3D Sagittal Magnetization Prepared Rapid Acquisition Gradient-Recalled Echo (MPRAGE) sequence. Repetition time (TR) was ≈2300 ms, echo time (TE) ≈3 ms, and inversion time (TI) = 900 ms. Voxel dimensions were ≈1.20 mm × 1.015 mm × 1.015 mm. All images were acquired using 8 channel head array receiver coils.

Area of acquisition: Whole-brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	SPM12 and in-house Matlab scripts
Normalization	Non-linear
Normalization template	Mayo Clinic Adult Lifespan Template (MCALT; https://www.nitrc.org/projects/mcalt/)
Noise and artifact removal	No procedure for artifact and structured noise removal was used.
Volume censoring	Not applicable.

Statistical modeling & inference

Model type and settings	Mass univariate analysis was performed at voxel level.
Effect(s) tested	MRI correlated of age and brain age gap were explored.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	Voxel-wise analysis was performed.
Correction	FDR-correction was used for multiple comparisons with a cluster size of at least 100 adjacent voxels as implemented in AFNI.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis	This study used a deep learning regression model. Mean absolute error and Spearman's correlation (between the observed and true values) metrics were evaluated for the regression model. Voxel-wise linear model was used to predict MRI intensity or FDG-PET uptake using brain age gap as a regressor. Each individual's chronological age was specified as nuisance covariance