Article

Morphological and genetic decoding shows heterogeneous patterns of brain aging in chronic musculoskeletal pain

Lei Zhao[1,2,9], Jiao Liu[3,4,9], Wenhui Zhao[1,2], Jie Chen[1,2], Jicong Fan[5], Tian Ge[6,7,8] & Yiheng Tu [1,2]✉

Chronic musculoskeletal pain (CMP), a prevalent and heterogeneous condition characterized by persistent pain in various body parts, is a leading cause of disability worldwide and greatly affects a patient's brain. Apart from experiencing pain, older adults with CMP also have accelerated cognitive decline and higher dementia risk with limited understanding of biological mechanism underlying the associations between CMP and dementia risk. A multiscale study to disentangle pathological brain aging from normal brain aging may reveal the underlying mechanisms. Using large-scale, cross-sectional and longitudinal cohorts ($N = 9,344$), we have developed an MRI-based brain age model ($N = 6,725$) to evaluate the difference between brain age and chronological age, termed 'predicted age difference' (PAD), across several common types of CMP ($N = 2,427$). Our study unveils significantly increased PAD in knee osteoarthritis (KOA) cohorts versus healthy controls, and validates it in an independent dataset ($N = 192$), suggesting a pattern of brain-aging acceleration in KOA. This acceleration was contributed by the hippocampus in both datasets and predicted memory decline and dementia incidents during follow-up. The *SLC39A8* gene showed pleiotropy between brain-aging accelerations and KOA and exhibited spatially transcriptional associations with the regional contributions to brain-aging accelerations. The genes exhibiting spatially strong transcriptional associations with regional contributions were highly expressed in microglial cells and astrocytes, and were mainly enriched in synaptic structure and neurodevelopment. These findings highlight a heterogeneous pattern of brain aging in CMP and reveal a heritable morphological pattern that links brain-aging acceleration to cognitive decline and an elevated risk of dementia in KOA.

Elderly individuals disproportionately suffer from chronic musculoskeletal pain (CMP), a major source of disability that affects more than 40% of the world's population[1]. Increasing evidence suggests that, apart from persistent pain, older adults with CMP are also subject to accelerated cognitive decline and an increased probability of dementia[2–5].

During the aging process, cognitive decline and increased dementia incidents are associated with the accumulation of impairments in the brain, at scales ranging from the molecular genetic level to cellular and morphological levels[6–8]. However, the accumulation of biological impairments has substantial heterogeneity in individuals with different

lifestyles, health conditions, environmental factors and genetic risk factors, and this modulates the rate of aging[9,10]. Thus, characterizing biological brain aging rather than chronological aging will have substantial implications in terms of explaining variations between cohorts and individuals regarding memory decline and dementia risks with CMP.

The brain plays a pivotal role in regulating mental health in relation to cognition, emotion and behavior[11]. The anatomical structure of the brain constrains its functional organization and thus offers insights into mental health[12]. The structure of the brain changes constantly with increasing chronological age. These changes mostly reflect the normal aging process, but they can also be modulated by pathological conditions and genetic predisposition[13–15]. Age prediction using magnetic resonance imaging (MRI) data and machine-learning techniques could provide a robust estimation of an individual's 'brain age' by assessing the age of brain tissues from a normative lifespan trajectory[16]. Studies suggest that the deviation of an individual's brain age from chronological age is associated with cognitive function[17,18]. The rate of brain aging can be accelerated by some disease exposures that have critical roles in the emergence and development of dementia[19]. For example, a higher brain age relative to chronological age has been found in patients with schizophrenia[20], depression[21] and alcohol dependence[22]. These studies suggest that brain age may hold prognostic value, potentially predicting cognitive decline and dementia risks by assessing the individual differences in the interaction of brain aging and CMP.

CMP is heterogeneous in etiology, spanning a wide range of genetic[23], environmental[24] and biological factors[25]. This complexity suggests that brain-aging trajectories in different CMP conditions may exhibit substantial heterogeneity. For example, chronic knee or hip pain, often attributed to knee osteoarthritis (KOA) and hip osteoarthritis, is frequently accompanied by persistent low-grade inflammation, which has been widely acknowledged as a hallmark of biological aging[26]. There is also evidence supporting the association of osteoarthritis with accelerated memory decline and an elevated risk of dementia[27,28]. A recent study has shown that osteoarthritis can accelerate the accumulation of amyloid-β and tau proteins in the brain, which are key pathological events associated with the development of Alzheimer's disease[29], highlighting the possibility that individuals with osteoarthritis may be at a higher risk of experiencing the acceleration of brain aging.

A few studies have assessed brain-aging acceleration in CMP[30,31], but several issues remain to be addressed (Fig. 1a). First, studies so far have been small in scale and have focused on a single type of CMP, hindering assessment of the clinical and mechanistic distinctions. It is unclear whether common types of CMP share a general pattern or follow distinct patterns of brain-aging trajectory. Second, a longitudinal investigation is warranted to uncover the potential of brain age as a prognostic biomarker for cognitive decline and dementia risk in patients with CMP. Third, the genetic underpinnings of brain aging in CMP are not understood, although genome-wide association studies (GWASs) have identified risk variants linked to CMP and brain-aging acceleration, respectively[32,33]. The recent construction of a brain-wide gene expression atlas has made it possible to connect macroscale spatial compositions of brain aging with spatial variations in microscale gene expressions[34], providing an alternate avenue to explore the molecular genetic basis of brain aging in patients with CMP.

To address these issues, we first trained an elastic net regression model using a cohort of healthy participants (training set, $N = 5,202$) from the UK Biobank (UKB), utilizing their chronological age as the label and structural MRI data as the feature. We refer to this model as the 'brain age model' because its purpose is to predict the age of an individual's brain. The generalization of this model was evaluated by testing it on a separate cohort of healthy participants (hold-out set, $N = 1,523$) from UKB. We then applied this model to several common types of CMP in dataset 1 ($N = 2,427$; from UKB) to assess which type of CMP accelerates brain aging, then validated the findings on dataset 2

($N = 192$; from the local community). Patients in dataset 2 were invited to re-evaluate their cognitive scores, assess their dementia risk, and receive a rescanning after five years. This longitudinal design enabled us to associate brain-aging acceleration at the baseline session with the cognitive decline at the follow-up session. Finally, we examined molecular genetic mechanisms of brain-aging acceleration using GWAS summary statistics, gene transcriptional profiles of the brain and gene markers specific to brain cell types.

Results

Brain age model
Using a training set ($N = 5,202$) from UKB, we trained the brain age model to fit the relationship between the pattern of whole-brain gray matter volume (GMV) and chronological age in healthy individuals. This fitting provides a predefined reference that informs a supposed position (that is, corresponding chronological age) in the healthy aging trajectory for a given brain. We then tested the performance of the brain age model in an independent hold-out set ($N = 1,523$) from UKB. The results showed that the brain age predicted by our brain age model closely matched the individual's chronological age (Pearson's $r = 0.928$, $P < 0.001$, confidence interval (CI) = (0.921, 0.935); mean absolute error (MAE) = 2.367) (Fig. 2a), suggesting that brain age effectively reflects the biological age of the brain. To validate the stability and reliability of the brain age predicted by our model, we examined the scan–rescan consistency of this metric in individuals who had two MRI scans in the hold-out set ($N = 104$, the interval of the two scans ranged from two to six years, mean interval = 2.260 years). A strong correlation (Pearson's $r = 0.987$, $P < 0.001$, CI = (0.981, 0.991)) was found between brain age in the first and second scans (Fig. 2b). The change of brain age between the two scans was significantly equivalent ($P < 0.001$, equivalence testing) to the change in chronological age (mean change of brain age = 2.322 years, mean interval of two scans = 2.260 years) (Fig. 2b). Overall, our brain age model provided an accurate and reliable estimation of an individual's brain age.

Heterogeneous patterns of brain aging in CMP
To investigate which types of CMP deviate from the normal trajectory of brain aging, we applied the established brain age model to individuals with CMP taken from UKB (dataset 1). Each individual in dataset 1 ($N = 2,427$) was assigned to one of four cohorts according to the single painful site relevant to them: chronic knee pain ($N = 982$), chronic back pain ($N = 591$), chronic neck pain ($N = 528$) and chronic hip pain ($N = 326$) cohorts. Consistent with a previous brain age study on multiple disease groups[32], for each CMP cohort we randomly selected sex- and age-matched healthy controls (HCs) of equal number from the hold-out set. For each individual, we estimated a predicted age difference (PAD), which was the deviation between an individual's brain age and their chronological age. This index provided an individual with a quantifiable evaluation of the level of brain-aging acceleration[16]. Next, between-group PAD differences in each CMP cohort were examined, relative to their respective HC.

Figure 2c shows the estimated PAD and statistical results in dataset 1. Compared with HC, the chronic knee pain cohort showed significantly higher PAD (Cohen's $d = 0.130$, false discovery rate (FDR) $q = 0.027$), but no significant alterations were observed in the other CMP cohorts (FDR $q > 0.05$). To examine whether increased PAD in the chronic knee pain cohort was dominated by KOA, the most common disease leading to chronic knee pain in older adults, we further subdivided the chronic knee pain cohort into two subgroups (that is, with ($N = 161$) and without ($N = 821$) KOA). Significantly higher PAD ($d = 0.437$, $P < 0.001$) was found in the KOA cohort, but not in those without KOA ($P > 0.05$), compared with sex- and age-matched HCs of equal size. We also performed several additional analyses to demonstrate that brain-aging acceleration in KOA was not (1) merely an exemplification of arthritis, (2) due to reduced physical activity and (3) dominated by
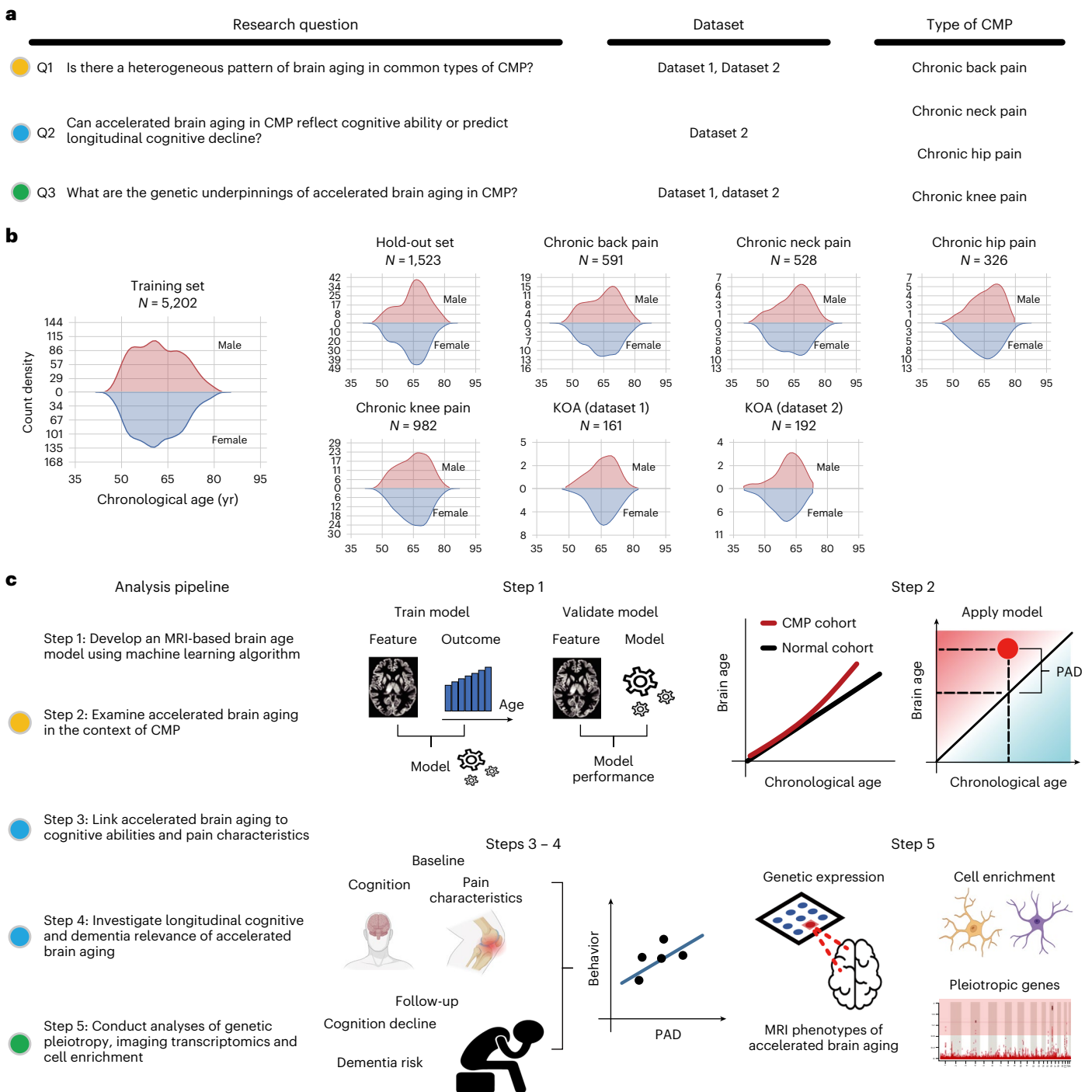
a, Research question | Dataset | Type of CMP

b

c, Analysis pipeline

Fig. 1 | Overview of research questions, participants and analysis pipeline.
a, The present study aims to answer three research questions (Q1–Q3) using two independent datasets (datasets 1 and 2) consisting of several common types of CMP. **b**, Age distributions of the training set, hold-out set and several cohorts with CMP. **c**, Overview of the five steps to address the three research questions. We have developed and validated an MRI-based brain age model using training and hold-out sets (step 1) and apply it to examine distinct brain-aging trajectories in several cohorts with CMP (step 2). We also investigate the relationships between brain-aging acceleration and cognitive function, dementia risk and pain characteristics (steps 3 and 4). Genetic analyses are employed to explore the genetic underpinnings of brain-aging acceleration (step 5). PAD, predicted age difference.

specific sex or (4) caused by medications, comorbidities and levels of education (Supplementary Figs. 1–4 and Supplementary Methods 1–4). In addition to gray-matter structures, white-matter structures and the functional activity of gray matter also showed substantial age-related changes[6,35]. Accordingly, we also constructed two brain age models based on the structural connectivity (SC) of brain white matter and the functional connectivity (FC) of brain gray matter and applied them to evaluate the brain-aging acceleration of CMP. Consistent with the findings for gray-matter structures, PAD estimated by structural and functional connectivity was significantly higher (SC: $d = 0.299$, FDR $q = 0.013$; FC: $d = 0.221$, FDR $q = 0.038$) in the KOA cohort compared to HCs (Supplementary Figs. 5 and 6 and Supplementary Method 5). It is worth noting that although KOA had the largest effect size (Cohen's d) in brain-aging acceleration estimated by FC, we observed that patients
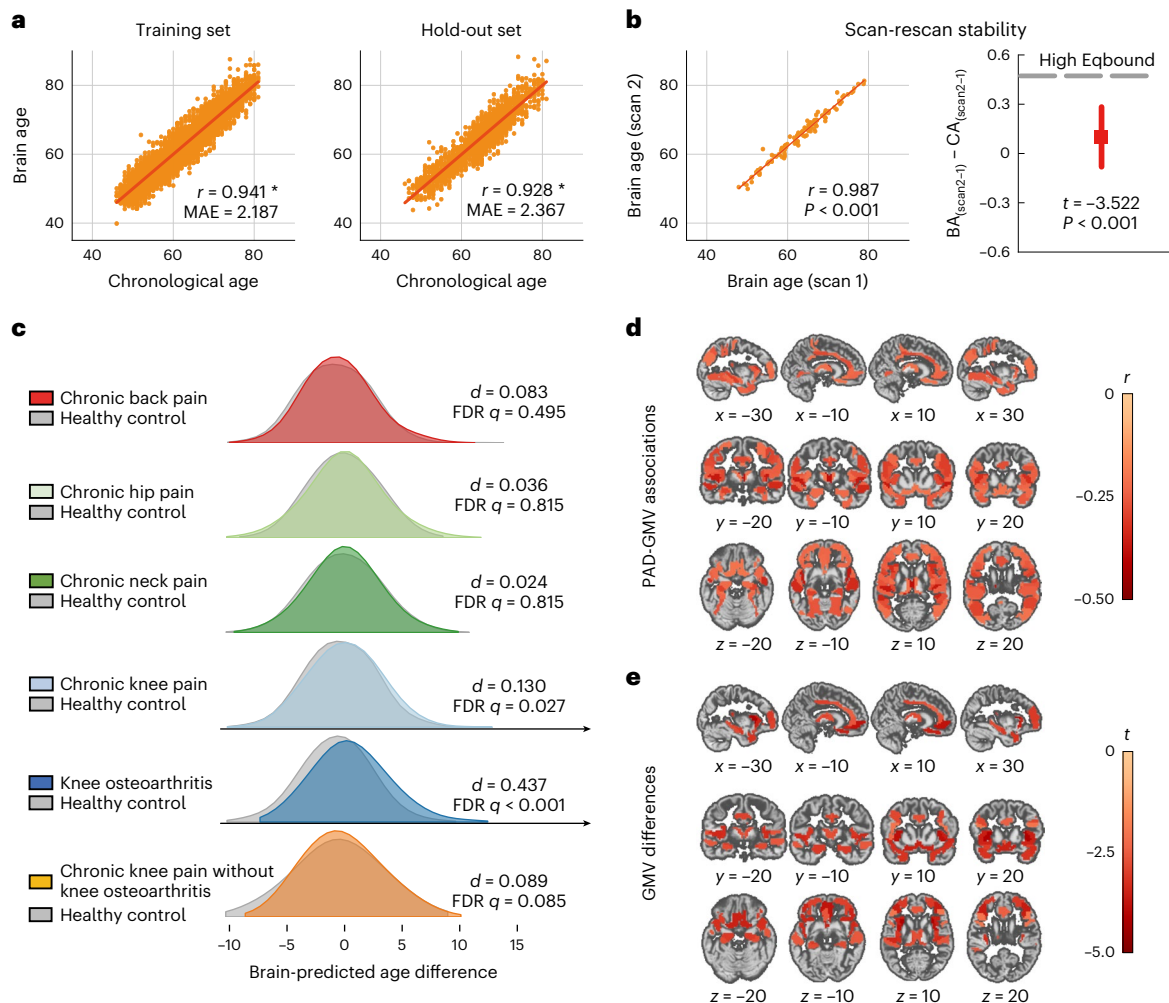
Fig. 2 | Training a brain age model and applying it to cohorts with CMP in dataset 1. a, Plots showing significant associations ($P < 0.001$, two-tailed Pearson correlation) between predicted brain age and chronological age in the training set (left, $N = 5,202$) and hold-out set (right, $N = 1,523$). **b**, Assessment of the stability of the brain age model across two scans ($N = 104$). Predicted brain age showed a significant association between the first and second scans (left, $P < 0.001$, two-tailed Pearson correlation). Significant equivalence was found between the longitudinal change of brain age and chronological age within a mean scan interval of 2.260 years (right, $P < 0.001$, one-tailed equivalence testing). The gray dashed line represents the upper equivalence bounds and the whiskers represent 95% CIs. **c**, PAD was significantly increased (two-tailed two-sample t-test) in cohorts with chronic knee pain and KOA compared to their respective HCs. **d**, Spatial brain patterns where the colored brain areas show significant associations (Bonferroni $P < 0.05$, two-tailed Pearson correlation) between GMV and PAD across the KOA and HC groups. **e**, Spatial brain patterns where the colored brain areas show significant associations both between GMV and PAD and difference in GMV ($P < 0.05$, two-tailed two-sample t-test) between the KOA and HC groups. *$P < 0.05$. d, Cohen's d; BA, brain age; CA, chronological age; High Eqbound, upper equivalent boundary.

with hip pain and other knee pain had higher PAD than controls. Given the highly fluctuating nature of FC, we believe that findings from structural MRI (that is, regarding gray matter volume and SC) may provide a more reliable insight into brain-aging acceleration in CMP patients.

To demonstrate the reliability of the findings, we applied the brain age model to dataset 2, which consisted of 133 patients diagnosed with KOA and 59 matched HCs. Figure 3a displays the estimated PAD and statistical results from dataset 2. Consistent with dataset 1, significantly higher PAD was also found in the patients with KOA ($d = 0.454$, $P = 0.020$). We performed sensitivity analyses to confirm that brain-aging acceleration in KOA was not dependent on the representation of GMV features or the machine-learning algorithm employed in both datasets (Supplementary Fig. 1 and Supplementary Method 6). Taken together, we identified a replicable (increased PAD was found across two independent datasets) and specific pattern (only observed in KOA but not in other CMP or arthritis syndromes that we investigated) of brain-aging acceleration in KOA.

Although brain age is characterized by the pattern of whole-brain GMV, accelerations in diseased populations may be driven by some core regions. To investigate which brain areas contribute to brain-aging accelerations in KOA, we first examined the associations between the GMV of brain areas (parcellated by the Brainnetome Atlas) and PAD across the KOA and HC groups in dataset 1. As shown in Fig. 2d, PAD was significantly correlated with the GMV of a wide range of brain areas (Bonferroni-corrected $P < 0.05$). Among these brain areas, the hippocampus, thalamus, insula, orbitofrontal lobe (OFC), inferior frontal gyrus (IFG), middle frontal gyrus (MFG), middle temporal gyrus (MTG), superior temporal gyrus (STG) and cingulate gyrus exhibited significantly reduced ($P < 0.05$) GMV in the KOA group compared to the HC group (Fig. 2e). Despite the relatively small sample size, dataset 2 also revealed significant correlations with PAD (Bonferroni-corrected $P < 0.05$) and between-group differences ($P < 0.05$) in the GMV of the hippocampus (Supplementary Fig. 7 and Fig. 3b). Detailed information regarding the statistical results is presented in Supplementary Tables 1–4.
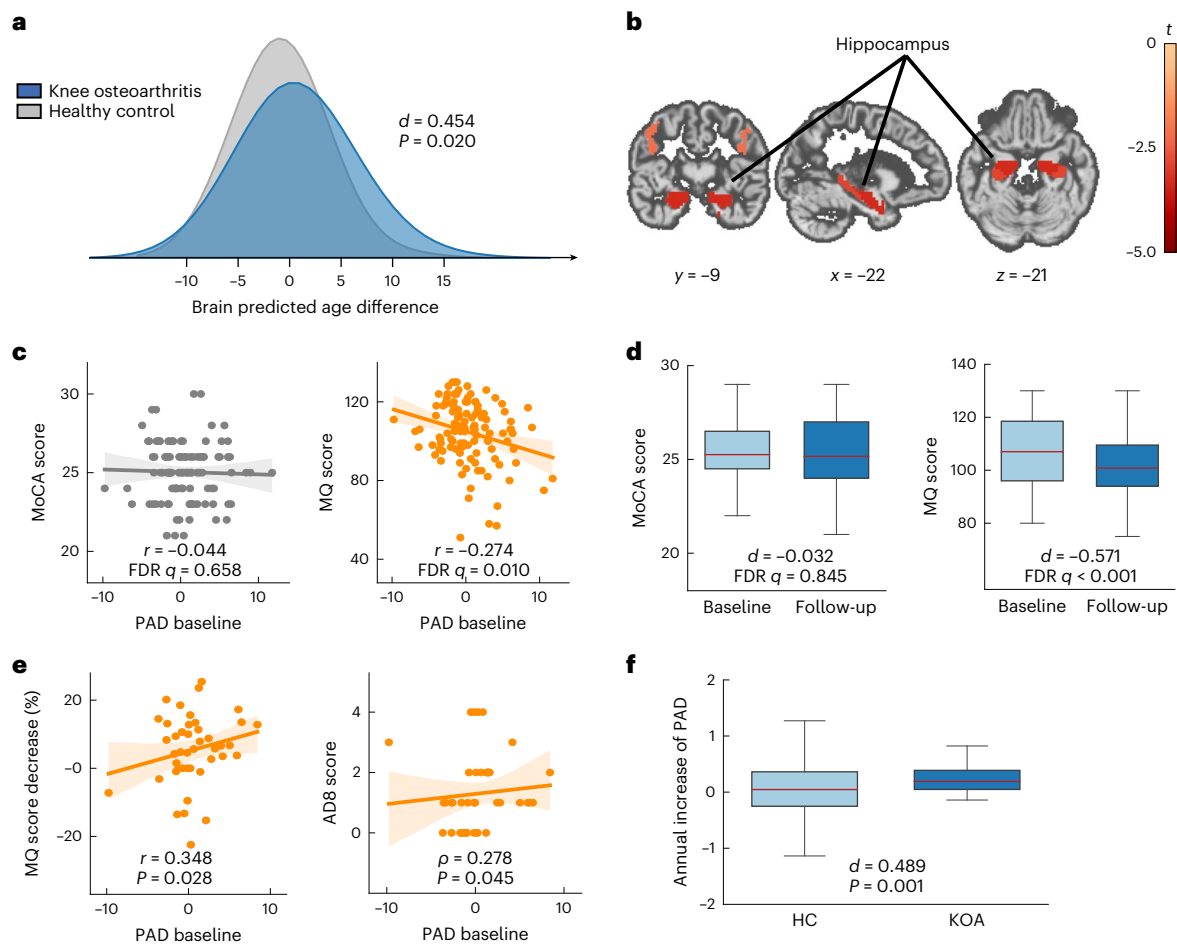
Fig. 3 | Brain-aging acceleration in KOA and its cognitive relevance in dataset 2. a, PAD was significantly increased ($P = 0.020$, two-tailed two-sample t-test) in cohorts with KOA compared to HCs. **b**, Spatial brain patterns in which the colored brain areas show both significant associations (Bonferroni $P < 0.05$, two-tailed Pearson correlation) between GMV and PAD and difference in GMV ($P < 0.05$, two-tailed two-sample t-test) between the KOA and HC groups. **c**, Scatter plots showing a significant association between baseline PAD and MQ scores in patients with KOA (right, FDR $q = 0.010$, two-tailed Pearson correlation), and no significant association between baseline PAD and MoCA scores in patients with KOA (left, FDR $q = 0.658$, two-tailed Pearson correlation). **d**, Box plots showing a significant difference between MQ scores measured in the baseline and follow-up sessions (right, FDR $q < 0.001$, two-tailed paired-sample t-test, $N = 43$), and no significant difference between MoCA scores measured in the baseline and

follow-up sessions (left, FDR $q = 0.845$, two-tailed paired-sample t-test, $N = 43$). **e**, Scatter plots showing a significant association between baseline PAD and reduction rate of MQ scores (left, $P = 0.028$, two-tailed Pearson correlation, $N = 43$) in a five-year follow-up and AD8 scores (right, $P = 0.045$, one-tailed Spearman correlation, $N = 41$) measured in the follow-up session in patients with KOA. **f**, Box plot showing a significant difference ($P = 0.001$, two-tailed two-sample t-test, $N_{KOA} = 43$, $N_{HC} = 104$) in annual increases of PAD between patients with KOA and HCs during follow-up. AD8, Ascertain Dementia 8; MoCA, Montreal Cognitive Function Assessment Scale; MQ, memory quotient. *$P < 0.05$. The results in the scatter plots are displayed as mean estimates (solid lines) with 95% CIs (shaded areas). The results in box plots are displayed as the mean (bold red horizontal line), the first and third quartiles (lower and upper hinges) and 1.5× the interquartile range (whiskers).

PAD relates to cognitive function

We extended our investigation to examine the potential association between brain-aging acceleration and cognitive functions in patients with KOA. Given the observed associations between the GMV of the hippocampus and brain-aging acceleration, as well as the differences in the GMV of the hippocampus between the KOA and HC groups in both dataset 1 and dataset 2, we conducted a more specific examination of the relationship between brain-aging acceleration and memory function. To this end, we tested the associations between PAD and global cognitive and memory function measured by the Montreal Cognitive Function Assessment Scale (MoCA) and the Wechsler Memory Scale – Chinese Revision (WMS-CR), respectively, in dataset 2. As shown in Fig. 3c, PAD was significantly correlated with the memory quotient (MQ; WMS-CR total scores adjusted by age effects) (Pearson's $r = -0.274$, FDR $q = 0.010$) but not MoCA scores (FDR $q > 0.05$) in patients with KOA. To test whether brain-aging acceleration was driven by pain characteristics experienced by patients with KOA, we measured the knee injury and

osteoarthritis outcome scores (KOOS) of those patients. However, there was no significant correlation between any dimension of KOOS and PAD (FDR $q > 0.05$; Supplementary Fig. 8 and Supplementary Method 7). We also evaluated the impact of pain characteristics on the memory function of patients with KOA, yet found no significant correlation between any dimension of KOOS and MQ scores (FDR $q > 0.05$; Supplementary Fig. 9).

PAD predicts longitudinal memory decline

Although we observed that brain-aging acceleration was associated with memory function in patients with KOA at the baseline session, a longitudinal examination was essential to identify whether these patients could experience accelerated memory decline and whether this alteration could be predicted by PAD. In the present study, 43 patients in dataset 2 were invited to revisit and re-evaluate their cognitive and memory function after five years (follow-up session). Between the baseline and follow-up sessions, no significant changes in daily activity

or adverse events were reported by these patients. We found that their MQ scores were significantly reduced (baseline session, 107.00 ± 15.52; follow-up session, 100.81 ± 15.49; $d = -0.571$, FDR $q < 0.001$, paired-sample t-test) between the two sessions (Fig. 3d). It should be noted that the reduction in MQ scores was not significantly dependent on the patient's chronological age (Pearson's $r = 0.224$, $P = 0.149$). No significant difference (baseline session, 25.26 ± 1.66; follow-up session, 25.16 ± 2.88; $d = -0.032$, FDR $q > 0.05$, paired-sample t-test) was found in MoCA total scores between the two sessions.

Next, we investigated whether PAD at baseline could serve as a potential neural marker to predict memory decline in KOA. We found that the PAD at the baseline session showed a significant positive correlation (Pearson's $r = 0.348$, $P = 0.028$, FDR $q < 0.05$; Supplementary Fig. 3e) with the reduction rate (RR = (baseline − post)/baseline) of MQ in the five-year follow-up. The correlation was still significant (Pearson's $r = 0.375$, $P = 0.019$; Supplementary Fig. 3 and Supplementary Method 8) after adjusting for the effects of pain medications. Because memory decline is the most common first symptom and a hallmark of patients with dementia, we hypothesized that higher PAD would be related to the elevated risk of dementia. The Ascertain Dementia 8 (AD8) questionnaire, which includes eight questions to be answered 'yes' or 'no', has been proven to be highly sensitive and specific in detecting early manifestations of dementia. In the follow-up session, 41 patients completed the AD8 dementia screening interview. Consistent with our hypothesis, we observed a significant positive correlation (Spearman's $\rho = 0.278$, $P = 0.045$, FDR $q < 0.05$; Fig. 3e) between PAD at baseline and AD8 scores assessed in the follow-up session. After adjusting for the effects of pain medications, the correlation remained significant (Spearman's $\rho = 0.278$, $P = 0.046$; Supplementary Fig. 3 and Supplementary Method 8).

We conducted additional analyses to provide further evidence for the associations between brain-aging acceleration and memory decline, as well as the risks of dementia in KOA. These analyses were based on the individuals with KOA acquired from UKB, using more relaxed inclusion criteria than our primary analyses. The results revealed a significant positive correlation (Pearson's $r = 0.283$, $P = 0.038$) between PAD at baseline and the reduction rate of memory function during follow-up. Moreover, PAD at the baseline of those diagnosed with dementia during follow-up exceeded the 90th percentile threshold (3.990 years versus 3.934 years) within the KOA cohort. The top 10% of individuals with KOA with higher PAD exhibited a significantly elevated risk of dementia (OR = 25.297, $P = 0.004$, Fisher's exact test) when compared to the remaining individuals in the KOA group. Further details regarding the inclusion criteria, numeric memory test and dementia diagnosis are provided in Supplementary Method 9.

Longitudinal alterations of PAD
Subsequently, we investigated the longitudinal alterations of brain-aging acceleration in patients with KOA. For this analysis we included patients ($N = 43$) and HCs ($N = 104$) who had undergone a follow-up MRI. The results, illustrated in Fig. 3f, revealed a significantly higher annual increase of PAD (($PAD_{Post} - PAD_{Baseline}$)/the interval between the two scans) in the KOA group compared to the HC group ($d = 0.489$, $P = 0.001$, two-sample t-test) between two sessions. These findings suggest a worsening trend in brain-aging acceleration over time for the KOA group.

Pleiotropic gene between PAD and CMP
Structural changes of the human brain are regulated by gene expression throughout life[36]. Accordingly, we explored the molecular genetic mechanism underlying brain-aging acceleration in KOA. We first examined the genetic overlap between PAD and KOA. After gathering GWAS summary statistics for PAD ($N = 20,170$) and KOA (case = 24,955, control = 378,169)[32,33], we utilized a conjunctional

FDR (cFDR) method to identify single-nucleotide polymorphisms (SNPs) that were associated with both PAD and KOA. We found five SNPs (Fig. 4a; FDR $q < 0.05$) located in two pleiotropic genes, $SLC39A8$ (rs13107325) and $NFAT5$ (rs12447326, rs11643240, rs11075730 and rs6499237). We also performed cFDR analyses between PAD and the other three types of CMP (chronic back pain (case = 80,588, control = 36,816), chronic hip pain (case = 40,152, control = 11,364) and chronic neck pain (case = 72,887, control = 32,509)). The results showed that there was no pleiotropic gene between PAD and the other CMP conditions (FDR $q > 0.05$).

Gene $SLC39A8$ transcriptionally links with PAD
We further examined whether the pleiotropic genes between KOA and brain-aging acceleration affect the regional contribution to brain-aging acceleration in KOA using the post-mortem data of six healthy adult donors from the Allen Human Brain Atlas (AHBA). The regional contributions to brain-aging acceleration in KOA were defined as the KOA neuroimaging phenotypes (see Methods for details). The spatial correlations between KOA neuroimaging phenotypes and gene expression profiles (Fig. 4b) from AHBA were calculated. After correcting for spatial autocorrelation (SA), we found that the expression of gene $SLC39A8$ was negatively correlated with the KOA neuroimaging phenotypes in both datasets 1 (Spearman's $\rho = -0.329$, FDR $q_{SA} = 0.010$) and 2 (Spearman's $\rho = -0.365$, FDR $q_{SA} = 0.010$) (Fig. 4c). No significant correlation was found between the expression of gene $NFAT5$ and the KOA neuroimaging phenotypes (dataset 1: Spearman's $\rho = -0.292$, FDR $q_{SA} > 0.05$; dataset 2: Spearman's $\rho = -0.003$, FDR $q_{SA} > 0.05$). The specificity of gene $SLC39A8$ was defined as its percentile on the ranked gene list, which was obtained by ranking the spatial correlation coefficients (that is, Spearman's ρ) on the whole gene sets (15,980 genes) averaged across two datasets. On the ranked gene list, the spatial correlation coefficient of gene $SLC39A8$ was located in the top 1% (Fig. 4d). Next, we replicated the analyses within the cortical and subcortical regions separately. Across the subcortical regions ($N = 18$), the expression of gene $SLC39A8$ was negatively correlated with the KOA neuroimaging phenotypes in both datasets 1 (Spearman's $\rho = -0.674$, $p_{SA} = 0.001$) and 2 (Spearman's $\rho = -0.649$, $p_{SA} = 0.003$) (Fig. 4c). No significant correlation was found across the cortical regions ($N = 105$) in either dataset ($p_{SA} > 0.05$; Fig. 4c), although the same trend of negative correlation was observed. Considering that the hippocampus contributes to brain-aging acceleration of KOA in both datasets, we also examined the effects of gene $SLC39A8$ on the hippocampal neuroimaging phenotypes of KOA. Across tissue samples within the hippocampus, we also observed a significant correlation between the expression of gene $SLC39A8$ and the KOA neuroimaging phenotypes in both datasets 1 (Spearman's $\rho = -0.319$, $p_{SA} = 0.048$) and 2 (Spearman's $\rho = -0.569$, $p_{SA} = 0.002$) (Fig. 4e).

PAD-related GO and cell-type enrichment
The Human Protein Atlas (https://www.proteinatlas.org) provides information on gene expression levels in the different cell types of brain tissue. In this atlas, gene $SLC39A8$ has the highest expression level in two glial cell types, microglial cells (Mic) and astrocytes (Ast) (Supplementary Fig. 10). Glial cells have been recognized as modulators of the neuronal environment and leaders in the progression of neurodegenerative diseases[37]. This drove us to identify whether glial cells preferentially express the genes with strong correlations with KOA neuroimaging phenotypes. Based on the cell-specific gene markers of eight classes of brain cell (Ast, endothelial cells (End), excitatory neurons (Ex), interneurons (In), Mic, oligodendrocytes (Oli), pericytes (Per) and oligodendrocyte precursor cells (OPC)) collected from a previous study[38], we performed cell-enrichment analyses on the ranked gene list (Fig. 5a) using a fast-preranked gene set enrichment analysis (FGSEA) method. All eight cell classes showed significant enrichment (FDR $q < 0.05$; Fig. 5b) on the top or bottom of the ranked gene list (Supplementary Appendix 1). Among the eight cell classes, Ast had the highest
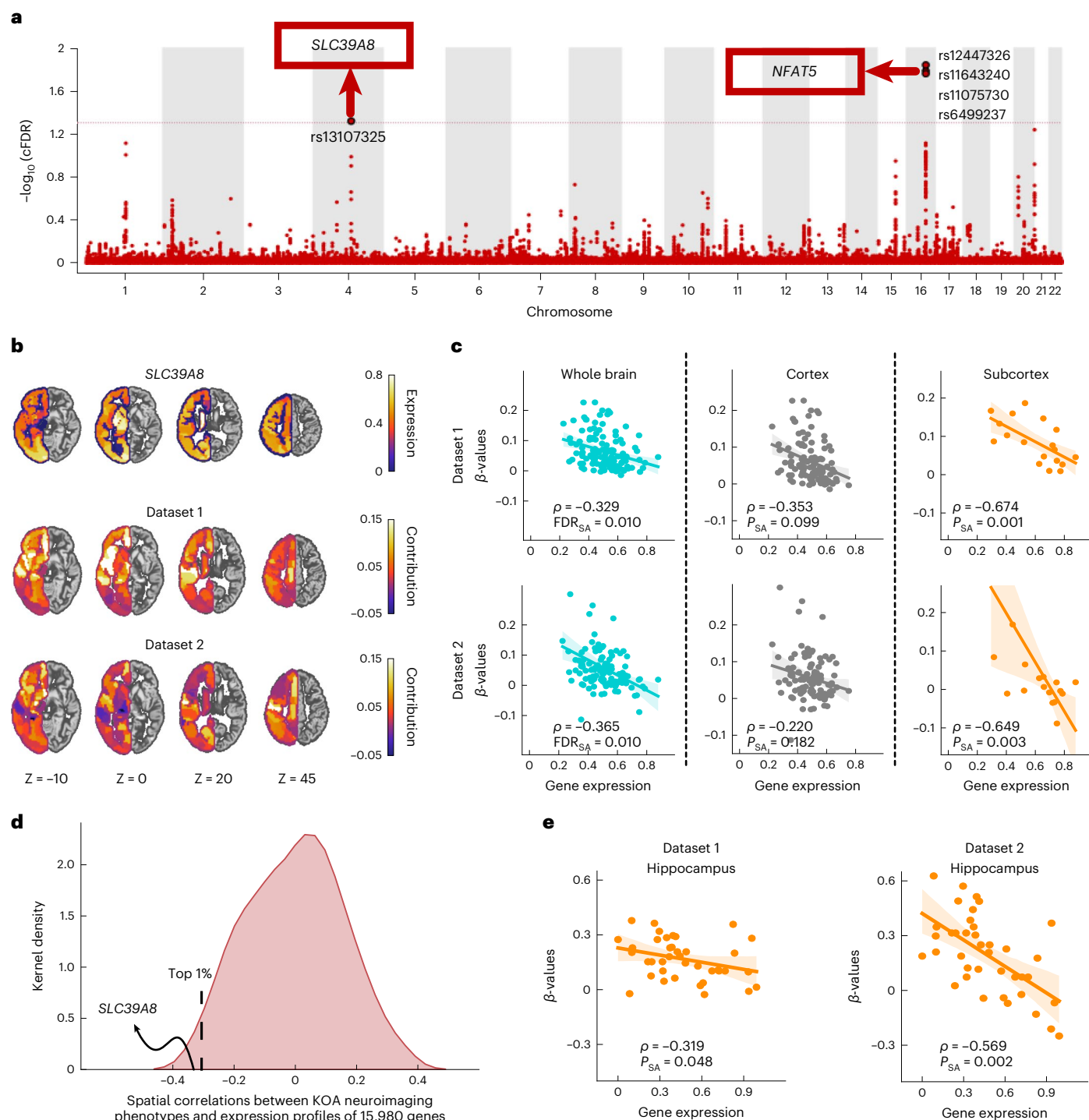
Fig. 4 | Pleiotropic genes between KOA and brain-aging acceleration.
a, Manhattan plot displaying five significant SNPs located in two pleiotropic genes in the cFDR analysis. The x axis shows chromosomal position and the y axis the significance ($-\log_{10}$FDR q; two-tailed). The red line marks the predefined cFDR significance level (FDR $q = 0.05$). **b**, Spatial maps of gene *SLC39A8* expression (top), KOA neuroimaging phenotypes in dataset 1 (middle) and KOA neuroimaging phenotypes in dataset 2 (bottom). **c**, Scatter plots showing the spatial association (two-tailed Spearman correlation) between gene *SLC39A8* expression and KOA neuroimaging phenotypes across the whole brain (left), cortical regions (middle) and subcortical regions (right) in dataset

1 (top) and dataset 2 (bottom). **d**, Distribution of spatial associations (two-tailed Spearman correlation) between KOA neuroimaging phenotypes and gene expression profiles across the brain. The dashed line indicates the top 1% negative-correlation genes among all 15,980 genes. The spatial association of gene *SLC39A8* is located in the top 1% among all 15,980 genes. **e**, Scatter plots showing the spatial association (two-tailed Spearman correlation) between the expression of gene *SLC39A8* and KOA neuroimaging phenotypes within the hippocampus (the contribution of 39 tissue samples located in the hippocampus to PAD) in dataset 1 (left) and dataset 2 (right). The results in the scatter plots are displayed as mean estimates (solid lines) with 95% CIs (shaded areas).

ratio of core genes and normalized enrichment score (NES), and Mic had the second highest ratio of core genes and NES. These findings can be replicated (Supplementary Fig. 11) based on the gene markers (Supplementary Appendix 2) derived from another study[39].

We identified two sets of Gene Ontology (GO) enrichment terms (Fig. 5c) that were significantly (FDR $q < 0.01$) enriched at the top decile of the ranked gene list. The identified biological process terms were mainly related to neurodevelopment, including the development of neuron (GO:0048666, FDR $q < 0.001$), neuron projection (GO:0031175, FDR $q < 0.001$) and central nervous system (GO:0007417, FDR $q < 0.001$). The cellular component terms were mainly involved in the synapse (GO:0045202, FDR $q < 0.001$), postsynapse (GO:0098794, FDR $q < 0.001$), presynapse (GO:0098793, FDR $q < 0.001$), axon (GO:0030424, FDR $q < 0.001$) and dendrite (GO:0030425, FDR $q < 0.001$). No significant molecular function term was found (FDR $q > 0.01$). We did not find a significant GO enrichment term at the bottom decile of the ranked gene list (FDR $q > 0.01$). A full list of GO enrichment terms is provided in Supplementary Appendix 3.

Discussion

In this study we have developed a brain age model to estimate brain-aging acceleration based on the morphological structures of the brain. Among several cohorts with different types of CMP, brain-aging acceleration characterized by increased PAD was only found in individuals with KOA. Importantly, this brain-aging acceleration in KOA was also validated in an independent dataset and showed continuous aggravation over time. The hippocampus, thalamus, insula, OFC, IFG, MFG, MTG, STG and cingulate gyrus were identified as key brain regions contributing to this acceleration. Furthermore, brain-aging acceleration at baseline not only showed a correlation with baseline memory function, but it also predicted future memory decline and potential dementia risk. The *SLC39A8* and *NFAT5* genes exhibited pleiotropy between KOA and brain-aging acceleration, and the spatial expression patterns of gene *SLC39A8* influenced the regional contributions to brain-aging acceleration in KOA. Additionally, genes preferentially expressed in Mic and Ast and those involved in synaptic structure and neurodevelopment showed the strong transcriptional associations with the regional contributions to brain-aging accelerations in KOA.

The MRI-based brain age model holds great promise for detecting disease risk[40], and the performance of this model benefits from large training samples[41]. In this study we trained a model based on more than 5,000 healthy samples from a publicly available data source that allows a more accurate estimation of brain age[42]. Aging-associated alterations in the brain usually exhibit sex-specific characteristics[6]. Accordingly, we matched the age distribution between males and females in the training set to capture a sex-neutral brain-aging trajectory. After training, we tested the trained model on a hold-out set and showed that the predicted brain age closely matched an individual's chronological age. In addition, using data from UKB repeat visits, we showed highly consistent brain-age estimates within participants across two MRI scans, suggesting high scan–rescan stability of our brain model. Overall, our trained model provides a robust and reliable estimation of an individual's brain age.

Applying the brain age model to several common types of CMP, we found brain-aging acceleration only in the chronic knee pain cohorts, suggesting heterogeneous patterns of brain aging in CMP. Subgroup analysis in the KOA cohort across two datasets revealed a replicable brain-aging acceleration characterized by increased PAD. Among the brain areas playing a key role in brain-aging acceleration in KOA, the contribution of the hippocampus was identified across two datasets. The hippocampus is critically involved in several cognitive domains, particularly memory encoding and consolidation[43]. It is also related to cognitive impairments in neurodegenerative diseases and is especially vulnerable in the early stages of Alzheimer's disease[44]. A contribution of the thalamus to brain-aging acceleration in KOA was found in dataset 1.

The thalamus is densely connected with the hippocampus and plays a central role in memory retrieval by modulating hippocampal–thalamic–cortical networks[45]. Volume losses of the hippocampus and thalamus can be found in patients with Alzheimer's disease, even in the mild stage, and are correlated with cognitive impairment[46]. In line with these findings, the brain regions associated with cognitive processing and memory function underlie brain-aging acceleration in KOA.

Our longitudinal investigation revealed that the memory function (measured by age-adjusted MQ) of patients with KOA was reduced after five years, but no reduction was observed in global cognitive function (measured by MoCA). This result implies that memory function might be one of the most vulnerable or earliest impaired cognitive domains in patients with KOA. Progressive memory decline is a hallmark of dementia, but it is hard for patients to be aware of this subjectively[47]. Morphological alterations in the brain usually precede apparent cognitive impairments, making them potentially sensitive to the accumulation of biological impairments for dementia[48,49]. PAD is sensitive to integrating and quantifying subtle but spatially widespread structural abnormalities associated with cognitive deficits[17]. Here we linked the pattern of morphological abnormalities at baseline and longitudinal memory decline, demonstrating that higher brain-aging acceleration at baseline can predict more considerable memory decline in patients with KOA. Moreover, a higher PAD in such patients was associated with a higher risk of dementia estimated by AD8, a sensitive screening tool with reliable discriminative power for early-stage dementia[50]. In line with this discovery, our supplementary analyses under relaxed inclusion criteria revealed that individuals in the KOA cohort with PAD in the top 10% exhibited a higher risk of dementia than the remaining individuals. Overall, these findings highlight the potential of brain-aging acceleration as a prognostic biomarker for cognitive decline and dementia in patients with KOA. Furthermore, between the MRI scans at the baseline and follow-up sessions, PAD was not constant, but instead further increased in patients with KOA, emphasizing the necessity to uncover the mechanism underlying brain-aging acceleration to prevent further cognitive impairments.

In the present study, increased PAD in patients with KOA was not explained by pain characteristics or physical activity. We thus sought to examine whether there is any gene underlying brain-aging acceleration in KOA. Consistent with neuroimaging findings, we found shared risk genes between PAD and KOA, but not the other types of CMP, providing genetic evidence for the heterogeneous patterns of brain aging in CMP. The shared risk gene *SLC39A8* encodes a transmembrane protein that plays critical functions in the onset of inflammation and is responsible for transporting several divalent cations related to synaptic plasticity and cognitive deficits[51]. Using imaging transcriptomics analysis, we consistently identified spatial transcriptional associations between gene *SLC39A8* and subcortical KOA neuroimaging phenotypes across two datasets. This result is consistent with previous research showing that the missense mutation of rs13107325 in gene *SLC39A8* is associated with structural variations in the subcortex[52]. In the brain, gene *SLC39A8* is mainly expressed in Mic and Ast. Ex vivo studies have indicated that the regional heterogeneity of genetic transcription, particularly within Mic and Ast, predetermines different regional susceptibilities to aging and neurodegenerative diseases[53,54], causing distinct changing trajectories in different cognitive domains[55]. Within the hippocampus, we also observed spatial transcriptional associations between gene *SLC39A8* and the KOA neuroimaging phenotypes. This finding is supported by earlier evidence that glial cells and the release of zinc might modulate hippocampal synaptic plasticity[56,57].

Based on enrichment analyses, we found that the gene markers of Mic and Ast and those associated with synaptic structure and neurodevelopment show strong spatial transcriptional associations with KOA neuroimaging phenotypes. Some studies have suggested that inflammation might modify brain structure in degenerative diseases via immune cells of the central nervous system (Mic and Ast)[58,59].
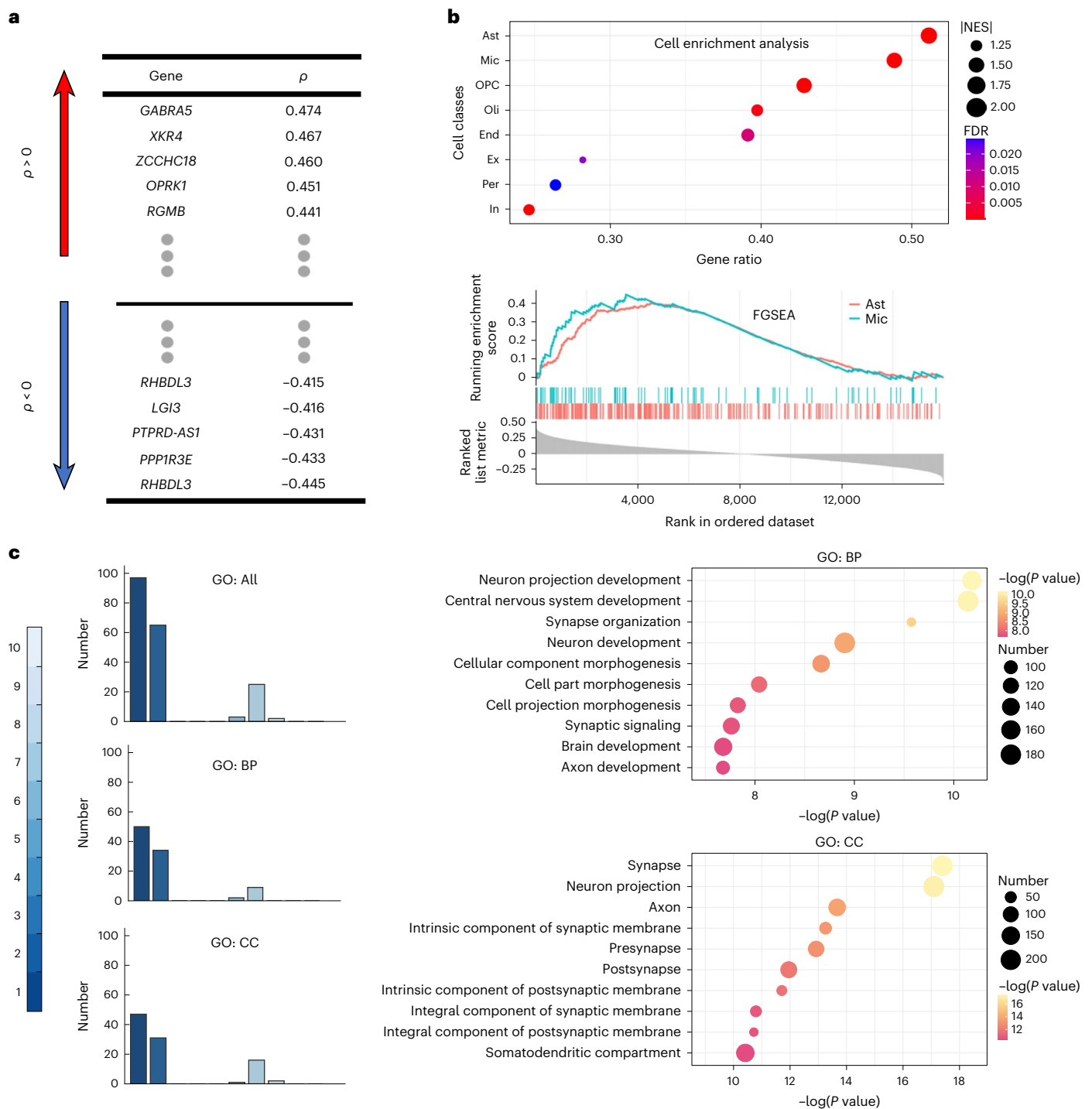
Fig. 5 | Enrichment analysis on genes with high correlations with KOA neuroimaging phenotypes. a, The ranked gene list was obtained by averaging the spatial correlation coefficients of KOA neuroimaging phenotypes on the whole gene sets (15,980 genes) across two datasets then ranking. **b**, Cell-enrichment analysis was performed on the ranked gene list. A total of eight sets of cell-specific gene markers were enriched on the top or bottom of the ranked gene list (top). The FGSEA enrichment plot shows that gene markers of Mic and Ast were enriched at the top of the ranked gene list (bottom). **c**, GO enrichment terms. The top gene deciles present the greatest number of enrichment terms in both BP and CC ontological categories (left). For the results of enrichment analysis on the top gene deciles, only the most significant ten enrichment terms are displayed in the figure (right). A one-tailed hypergeometric test was used for determining statistical significance (FDR-B&Y correction). |NES|, absolute value of the NES. BP, biological process; CC, cellular component.

Progression of KOA involves complex alterations in the inflammatory environment. The increased circulating concentrations of pro-inflammatory cytokines might alter brain structures via microglia and astrocytic function by means of abnormal synaptic pruning and the subsequent effect on GMV[60,61], especially in the hippocampus[62].

In addition, the interaction of the proinflammatory cytokine and glial cells might increase susceptibility to neurodegenerative diseases and cognitive impairments by producing dysregulation of synaptic plasticity and transmission[63,64]. Increased neuroinflammation in the thalamus of patients with KOA has been revealed by some neurometabolites

regarded as glial markers[65]. A positron emission tomography study using an in vivo marker of Mic and Ast has revealed a positive correlation between neuroinflammation in the brain and cognitive impairment score[66]. Taken together, biological processes involving glial cells and inflammation might connect KOA to brain-aging acceleration. We also found indications that inflammation may play a role in the heterogeneous pattern of brain aging in common types of chronic pain (Supplementary Fig. 12 and Supplementary Method 10). Our results show that, among the CMP groups examined, those suffering from chronic knee pain had the highest levels of C-reactive protein (CRP), an indicator of systemic inflammation, and the KOA group had higher CRP levels than the group with chronic knee pain. Moreover, the heterogeneous patterns of brain-aging acceleration were significantly correlated with variation in the magnitude of CRP level increases across different CMP groups.

Several limitations should be considered when interpreting our findings. First, the types of CMP were defined based on painful sites in the body because of the limitations of the dataset employed in the present study. This approach may not be optimal, as it does not always reflect the different underlying pathological mechanisms. Future investigations on PAD of CMP could integrate the types of CMP based on pathological similarity. Second, the CMP in UKB is mainly self-reported by individuals. Although some primary findings in UKB have been replicated in patients with KOA with a clinical diagnosis, the results of other types of CMP deserve further validation in diagnosed patients. Third, the sample size for the longitudinal analysis was relatively small, and the assessment of dementia risk was reliant on a questionnaire. Although the AD8 questionnaire has demonstrated notable sensitivity and specificity in identifying early signs of dementia, and our supplementary analyses under relaxed inclusion criteria supported the primary finding, further prospective studies incorporating a clinical dementia diagnosis and larger sample sizes are crucial for corroborating our findings. Finally, our findings point to the promise of brain-aging acceleration as a prognostic biomarker for dementia in KOA, but brain-aging acceleration alone cannot fully explain the accumulated biological impairments in the course of aging. To obtain a clinically implementable risk score for dementia, future works should combine brain age with other aging-related biological information, such as DNA methylation and telomere length.

In conclusion, we have trained a brain age model based on a large sample and identified specific brain-aging acceleration in individuals with KOA, contrasting with several common types of CMP across two independent datasets. This acceleration was primarily driven by the brain structures for cognitive processing and is related to longitudinal memory decline and dementia risk. Furthermore, we have demonstrated that *SLC39A8*—a gene highly expressed in glial cells—might be a key genetic factor underpinning this acceleration. Gene markers of Mic and Ast and those involved in synaptic structure and neurodevelopment were particularly strong transcriptional associates of the regional contributions to this acceleration. Together, we have demonstrated the heterogeneity of brain aging in CMP and identified a distinct heritable brain-aging acceleration pattern linking KOA to dementia by providing an integrative biological profile that connects specific genes, molecular processes and cell classes with morphological alterations.

Methods
Study design
The present study explored the patterns of brain aging in four common types of CMP (chronic knee pain, chronic back pain, chronic neck pain and chronic hip pain) and the cognitive relevance and molecular genetic basis of brain-aging acceleration by performing five steps of analyses (Fig. 1c) across several healthy and CMP cohorts (Fig. 1b). In step 1, we trained and validated a brain age model using a training set ($N = 5,202$) and a hold-out set ($N = 1,523$) consisting of healthy individuals from UKB. In step 2, we first applied the brain age model to six CMP

cohorts (the aforementioned four types of CMP and two subgroups of the chronic knee pain cohort) from UKB (dataset 1; $N = 2,427$) to estimate their distinct patterns of brain aging. Following identification of the CMP cohort that showed brain-aging acceleration, we validated the findings based on the same type of CMP cohort (that is, the KOA cohort in this study) and HCs from a locally collected dataset (dataset 2, $N = 192$). In steps 3 and 4, we further extended the findings based on cognitive and clinical measures acquired from dataset 2. Specifically, step 3 investigated the associations of brain-aging acceleration with global cognitive function, memory function and pain characteristics in patients with KOA from dataset 2. Step 4 included the KOA patients from dataset 2 who completed a five-year follow-up. The associations of their brain-aging acceleration at baseline with cognitive and memory decline, as well as dementia risk during follow-up, were investigated. In step 5, the molecular genetic basis of brain-aging acceleration was examined based on KOA neuroimaging phenotypes and publicly available genetic data.

Participants
The data used in the present study were acquired from two independent data sources: UKB and the Affiliated Rehabilitation Hospital of Fujian University of Traditional Chinese Medicine. This study was approved by the Human Research Ethics Committee at the Institute of Psychology of Chinese Academy of Sciences (ethical approval no. H21030). The study followed the STROBE (strengthening the reporting of observational studies in epidemiology) statement[67].

The present study included a subset of participants from UKB (application ID 71901) consisting of 6,725 healthy individuals and 2,427 individuals with CMP. UKB has been approved to obtain and disseminate data and samples by the North West Multi-Centre Research Ethics Committee (http://www.ukbiobank.ac.uk/ethics), and these ethical regulations cover the work in the present study. All participants voluntarily participated in the experiments carried out by UKB and provided written informed consent. To develop a brain age model, 6,725 healthy individuals were subdivided into a training set and a hold-out set. A proper splitting strategy for data is necessary to obtain a credible estimation for model performance[68]. According to a recent technical study on brain age[42], 5,000 samples are sufficient to provide good performance, with a further increase in the number of samples producing few improvements. Moreover, different age distributions between males and females may also cause potential bias for a brain age model because of differential brain developmental trajectories by sex[6,69]. In the present study, we randomly split 6,725 healthy individuals into a training set and a hold-out set 10,000 times, and determined an optimal split with the following criteria: (1) the number of samples in the training set was more than 5,000, (2) the age distribution between males and females was matched in both the training and hold-out sets and (3) the number of samples in the hold-out set was not less than any CMP cohort. The similarities between age distributions were estimated using Jensen–Shannon divergence, which is a common method for symmetrically measuring the similarity between two probability distributions[70]. In this procedure, 5,202 healthy (male = 2,394) individuals were assigned to the training set, and the remaining 1,523 healthy individuals (male = 701) were assigned to the hold-out set. Subsequently, 2,427 individuals who reported CMP only at one body site over three months via a touchscreen questionnaire (category ID 100048) were subdivided into four cohorts of dataset 1 according to their painful site: chronic knee pain ($N = 982$), chronic back pain ($N = 591$), chronic neck pain ($N = 528$) and chronic hip pain ($N = 326$). The chronic knee pain cohorts were further subdivided into the KOA cohort ($N = 161$) and the cohort without KOA ($N = 821$). Baseline T1-weighted MRI images were collected from all included participants. A total of 104 participants in the hold-out set completed a second MRI scan two to six years after the baseline scan. See Supplementary Fig. 13 and Supplementary Table 5 for detailed inclusion criteria and demographic information.

The data acquired from the Affiliated Rehabilitation Hospital of Fujian University of Traditional Chinese Medicine included 133 patients diagnosed with KOA and 59 sex- and age-matched HCs. The KOA patients were diagnosed in the left or right knee based on the diagnostic criteria of the American Rheumatism Association by an orthopedic physician. These patients achieved a grade 2–3 KOA on the Kellgren–Lawrence scale on a standing anterior–posterior X-ray. These 192 individuals were assigned to dataset 2. The present study obtained approval from the Medical Ethics Committee of the Affiliated Rehabilitation Hospital of Fujian University of Traditional Chinese Medicine and the second affiliated hospital of Fujian University of Traditional Chinese Medicine. Written informed consent was obtained from all participants, and registration was completed on the Clinical Trial Registry (http://www.chictr.org.cn/register.aspx; clinical trial registration no. ChiCTR-IOR-16009308). Baseline T1-weighted MRI images were collected from all included participants. Forty-three participants completed a repeat MRI scan and cognitive assessments in the five-year follow-up. Participants were compensated monetarily for each complete MRI session. See Supplementary Fig. 14 and Supplementary Table 6 for further descriptions of the inclusion criteria and demographic information.

Acquisition and preprocessing of imaging data

T1-weighted MRI images acquired from UKB are publicly available. Details about the image acquisition are available at the UKB website (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=2367) and in an open-source document (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf). T1-weighted MRI images in dataset 2 were acquired on a 3.0-T GE 750 scanner and eight-channel receiver head coil (General Electric) with the following parameters: slice thickness = 1 mm, flip angle = 15°, field of view (FOV) = 240 mm and 160 slices in acquisition.

T1-weighted MRI images were preprocessed using the Computational Anatomy Toolbox (CAT12, http://www.neuro.uni-jena.de/cat) with default settings, primarily including segment, normalization, modulation and smooth. In brief, denoised and bias-corrected T1-weighted MRI images were spatially registered to tissue probability maps provided by statistical parametric mapping (SPM) (http://www.fil.ion.ucl.ac.uk/spm) and segmented into gray matter, white matter and cerebrospinal fluid. After being spatially normalized into standard stereotactic Montreal Neurological Institute space, the modulated GMV, compensated for the effect of spatial normalization, was smoothed with a 4-mm full-width at half-maximum Gaussian kernel and resampled to 3-mm spatial resolution. The estimated intracranial volume (TIV) was calculated as the summation of the gray matter, white matter and cerebrospinal fluid volumes in native space.

Training and generalization of the brain age model

In the present study, the brain age model was constructed based on elastic net regression, a technique commonly used for brain age estimation[71]. In this model, the hyperparameter α was pre-set to 0.5 to balance the weight of L1-norm versus L2-norm regularizations. Regularization coefficient λ was tuned using a fivefold cross-validation on the training set, with 50 repetitions to realize different partitions on the dataset. The optimal λ determined by the cross-validation procedure was used to develop a final predictive model on the entire training set. Before training, GMV features were first standardized and transformed into features embedded in low-dimensional subspaces to speed up the training process and avoid overfitting. Specifically, each GMV feature was standardized (minus the mean μ and scaled to the unit variance σ) independently across training samples. Principal component analysis (PCA) was subsequently run on the standardized GMV features, and the top 500 components (explaining 68% of the total variance) were used as the feature inputs of the predictive model. Standardization and PCA were only performed on the training set, and the resulting

transformation parameters were stored and later applied to unseen test data. The performance of the brain age model was assessed on the hold-out set, which was independent of the model's training. We used the Pearson's correlation coefficient and MAE between predicted age and chronological age to measure the accuracy of the age estimation. The PCA and elastic net regression were run on MATLAB.

Calculation of PAD

After assessing the model performance, the established age predictive model was applied to preprocessed MRI data in datasets 1 and 2 to predict an individual's brain age. Next, an individual's PAD was computed by subtracting their chronological age from their brain age. To adjust for the confounding effect caused by sex and age, as frequently reported in previous studies on brain age, we constructed a regression model according to

$$\text{Brain age} = B_0 + B_1 \times \text{chronological age} + B_2 \times \text{sex} + \varepsilon \qquad (1)$$

as advised by a previous technical study[19], where B_0 is the intercept, B_1 and B_2 are the effects of chronological age and sex on brain age, respectively, and the residuals ε represent the corrected PAD controlling for sex and chronological age. The regression model was constructed only in the training set, and the resulting regression coefficients were used for bias removal in the test data as follows:

$$\text{Corrected PAD} = \text{brain age} - (B_0 + B_1 \times \text{chronological age} + B_2 \times \text{sex}) \qquad (2)$$

The corrected brain age was calculated by adding the chronological age and corrected PAD, as follows:

$$\text{Corrected brain age} = \text{chronological age} + \text{corrected PAD} \qquad (3)$$

In the present study, the corrected PAD and brain age were used to perform the following statistical analyses.

Cognitive evaluation

In dataset 2, cognitive evaluations were conducted using MoCA[72] and WMS-CR[73]. MoCA evaluates comprehensive cognitive ability, which is composed of seven different cognitive domains (visuospatial/executive functions, 5-min delayed verbal memory, verbal memory registration and learning, attention, naming, abstraction and orientation). WMS-CR measures seven different types of memory function (mental control, picture recognition, visual reproduction, associative learning, touch, comprehension memory and digit span). By adjusting for age effects from WMS-CR total scores, MQ was further calculated to measure the memory function corrected for age-related decline. We used MoCA total scores and MQ to assess the global cognitive function and memory function of patients with KOA, respectively. A higher total score of MoCA or MQ means better function. The AD8 questionnaire, a brief informant-based measure with eight questions, was used to evaluate dementia risk in the patients[74].

Genetic pleiotropy analysis

The cFDR method, inspired by the empirical Bayes statistical framework, is specifically designed to be applied to polygenic traits and disorders characterized by numerous subtle genetic effects. It combines GWAS summary statistics of two traits to identify pleiotropic loci that inform shared genetic risks, enabling the discovery of connections between brain structure and diseases[75], as well as between different diseases[76,77], and is tailored for scenarios characterized by dense elements[78,79]. In the present study we employed cFDR analysis to identify pleiotropic genes that were significantly associated with both KOA and brain-aging acceleration. We collected the GWAS summary statistics for KOA and brain-aging acceleration from previous studies[32,33] to perform

the cFDR analysis. We also investigated whether there are pleiotropic genes between brain-aging acceleration and the other types of CMP (chronic back pain, chronic hip pain and chronic neck pain)[80]. The cFDR analysis was run using the MATLAB-based pleiofdr toolbox (https://github.com/precimed/pleiofdr).

Preprocessing of gene expression data

Human gene expression data in the brain were acquired from AHBA, a publicly available transcriptional atlas provided by the Allen Institute[34]. A total of 3,702 spatially distributed tissue samples in AHBA were collected from the brains of six healthy adult donors without a history of psychiatric or neuropathological disorders. More details are available on the website of the Allen Institute (http://www.brain-map.org). Given that the human reference genome is constantly updated, reannotating probes to their corresponding genes is essential to obtain more accurate inferences. We completed probe-to-gene reannotation based on Allen probe sequences, provided using the Re-annotator toolkit (https://sourceforge.net/projects/reannotator). Following reannotation, we used the abagen toolbox (https://www.github.com/netneurolab/abagen)[81] to preprocess microarray data using the default procedures. Briefly, probes were first filtered to discriminate expression signals from background noise. Only the probe with the highest differential stability was retained for each gene, resulting in 15,981 genes with a single expression value. Then, tissue samples were assigned to brain regions parcellated by the Brainnetome Atlas (http://atlas.brainnetome.org). In total, 123 brain regions of the left hemisphere (105 cortical regions and 18 subcortical regions) were considered for the following analyses, but there were only two donors with tissue samples from the right hemisphere[34]. The between-sample and inter-donor variabilities were corrected by combining the robust sigmoid function and min–max function. In this procedure, one gene was discarded due to poor normalization. Finally, corrected genetic expression profiles were averaged across the six donors, resulting in a $123 \times 15,980$ expression matrix, with each column denoting the regional expression pattern of a gene.

Imaging transcriptomic analysis

The contribution of a brain voxel to brain-aging acceleration is affected by both the influence of GMV in that voxel on PAD and the impact of the presence of the disease on the GMV of that voxel. Therefore, we defined the contribution of a brain voxel to brain-aging acceleration as the product of the regression coefficients representing the association between group and GMV, and the association between GMV and PAD. These two regression coefficients are derived from two models, one with group and GMV as independent and dependent variables, respectively, and the other with GMV and PAD as independent and dependent variables, respectively. Both models incorporate chronological age, sex and TIV as covariates. The contributions of brain voxels to brain-aging acceleration in the whole brain were averaged into 123 brain regions in the left hemisphere parcellated by the Brainnetome Atlas, yielding region of interest (ROI)-wise neuroimaging phenotypes that present the contributions of 123 brain areas to brain-aging acceleration. Subsequently, the spatial correlation coefficients (Spearman's ρ) between the expression values of pleiotropic genes in AHBA and ROI-wise neuroimaging phenotypes were computed separately in datasets 1 and 2. In imaging transcriptomic studies, the false positive rate can be inflated by intrinsic expression or phenotype correlations among neighboring brain regions (that is, spatial autocorrelations)[82,83]. We corrected this spatial autocorrelation effect using the BrainSMASH package (https://brainsmash.readthedocs.io/en/latest/index.html), which introduces a spatially constrained null model[84]. To evaluate the specificity of a gene, spatial correlation coefficients (Spearman's ρ) for each AHBA gene were calculated, averaged across two datasets, and ranked. The target gene was considered to be specific if it was ranked within the top 5% of the ranked gene list (15,980 genes).

Enrichment analysis

ToppGene (https://toppgene.cchmc.org) was employed to identify biological enrichment terms of the genes highly correlated to neuroimaging phenotypes. Matching previous approaches[38,82], the aforementioned ranked gene list was first split into evenly sized gene deciles, then gene enrichment analyses were performed on the top decile. Three GO classes (biological process, cellular component and molecular function) were considered in the present study ($q < 0.01$, one-tailed hypergeometric test, FDR-B&Y correction). To perform cell-enrichment analysis, we acquired cell-specific transcriptional gene markers of 16 cell classes (Ast, End, Ex1, Ex3, Ex4, Ex5, Ex8, In1, In3, In4, In6, somatostatin, Mic, Oli, Per and OPC) from a previous study[38] that identified cell-specific gene markers based on transcriptional data in the frontal (BA 6/9/10) and visual cortex (BA17) obtained from the Gene Expression Omnibus[85]. In this study[38], differential expression within each cell class, in relation to all other classes, was calculated, and genes exhibiting significantly positive differential gene expression were defined as distinctive cell-specific gene markers of the respective cell class. Following a previous study that performed a clustering analysis on multiple cell classes[86], we further assigned Ex and In subtypes into Ex and In cell classes, respectively, resulting in eight canonical cell classes (Ast, End, Ex, In, Mic, Oli, Per and OPC). Based on the eight sets of cell-specific gene markers (Supplementary Appendix 4), FGSEA[87] was conducted to identify cell classes in which the gene markers were over-represented on the top or bottom of the ranked gene list. Single-cell FGSEA was run on the R-based cluster-Profiler package (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html). Two measures, the ratio of core genes and NES, were used to evaluate the degree of enrichment. The enrichment score was calculated by walking down the ranked gene list from the top. Through the ranked gene list, the enrichment score was increased when we encountered a gene belonging to the predefined gene set and decreased when it was not encountered[88]. Because the increase in the magnitude depends on the spatial correlation coefficients of the gene, a higher enrichment score would be obtained if gene markers were over-represented on the top or bottom of the ranked gene list. Enrichment scores were further transformed to NES by adjusting for variations in the gene set size of several cell classes. The ratio of core genes was calculated using a leading-edge analysis and represented how many genes directly contributed to the enrichment score[88]. Because cell-specific gene markers are sensitive to the acquisition methodology, analysis process or thresholding, we replicated FGSEA on the cell-specific gene markers (Supplementary Appendix 5) acquired from another study[39].

Statistical analysis

The performance of the brain age model was estimated using Pearson's correlation coefficient and MAE between the predicted age and chronological age. A higher Pearson's correlation coefficient and lower MAE indicate better predictive power of the brain age model.

The equivalence between the longitudinal change of the predicted age and chronological age in the hold-out set was examined using equivalence testing (one-tailed, $P < 0.05$), with 0.452 years as the upper equivalence bound (assuming that the error per year was no more than 0.2 years). Two-sample t-tests (two-tailed, $P < 0.05$) were used to perform between-group inference on PAD and annual increases of PAD during follow-up, with chronological age, sex and TIV included as covariates. The differences in cognitive measures between the baseline and follow-up sessions were estimated using a paired-sample t-test (two-tailed, $P < 0.05$). The relationships between PAD and cognitive measures and clinical scores were estimated using Pearson correlation analyses (two-tailed, $P < 0.05$) with chronological age, sex and TIV included as covariates. Considering that the AD8 score is a ranked variable and has strong non-normality (a positive-skewed distribution), we used Spearman correlation analyses (one-tailed, $P < 0.05$) to test the

relationship between PAD and AD8 scores with chronological age, sex and TIV included as covariates. Consistent with previous studies[39,89], Spearman correlation analyses (two-tailed, $P < 0.05$) were conducted to calculate the spatial correlation between the neuroimaging phenotypes and gene expression. The statistical results were corrected for multiple comparisons with FDR.

The associations between the GMV of brain areas and PAD were estimated using Pearson correlation analyses (two-tailed, $P < 0.05$) with chronological age, sex, TIV and disease groups included as covariates, followed by the Bonferroni multiple testing correction. Between-group inferences on GMV were performed using two-sample t-tests (two-tailed, $P < 0.05$), with chronological age, sex and TIV included as covariates.

The statistical significance of cell-enrichment analyses was obtained by permuting the gene symbols of the ranked gene list. The P value of the observed enrichment was calculated relative to the null distribution produced from permuted data and then adjusted to account for multiple testing corrections using the FDR procedure. To obtain a stable null distribution, we permuted the gene symbols 100,000 times.

Reporting summary

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

Data availability

The MRI data are from multiple sources. The data acquired from UKB are publicly available upon third-party authorization (https://www.ukbiobank.ac.uk/register-apply). The GWAS summary statistics of KOA were downloaded from https://www.ebi.ac.uk/gwas/publications/30664745. The GWAS summary statistics for chronic hip pain, chronic back pain and chronic neck pain were downloaded from https://gwas.mrcieu.ac.uk/datasets with the category codes as ukb-b-133, ukb-b-8463 and ukb-b-16118, respectively. Human brain gene expression data in AHBA can be downloaded from http://human.brain-map.org/static/download. Individual data used to estimate brain age in dataset 2 can be accessed via ScienceDB (https://doi.org/10.57760/sciencedb.psych.00120).

Code availability

The publicly available software for the analyses has been described in the Methods of our manuscript. Our custom analysis code and the developed brain age model can be accessed at https://github.com/tulab-brain/BrainAgeCP.

References

1. El-Tallawy, S. N. et al. Management of musculoskeletal pain: an update with emphasis on chronic musculoskeletal pain. *Pain Ther.* **10**, 181–209 (2021).
2. Moriarty, O., McGuire, B. E. & Finn, D. P. The effect of pain on cognitive function: a review of clinical and preclinical research. *Prog. Neurobiol.* **93**, 385–404 (2011).
3. Rouch, I. et al. Association between chronic pain and long-term cognitive decline in a population-based cohort of elderly participants. *Pain* **162**, 552–560 (2021).
4. Tian, J. et al. Association between chronic pain and risk of incident dementia: findings from a prospective cohort. *BMC Med.* **21**, 169 (2023).
5. Whitlock, E. L. et al. Association between persistent pain and memory decline and dementia in a longitudinal cohort of elders. *JAMA Intern. Med.* **177**, 1146–1153 (2017).
6. Bethlehem, R. A. I., Seidlitz, J. & White, S. R. Brain charts for the human lifespan. *Nature.* **604**, 525–533 (2022).
7. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
8. Xia, X. et al. Aging and Alzheimer's disease: comparison and associations from molecular to system level. *Aging Cell* **17**, e12802 (2018).
9. Ferrucci, L. & Kuchel, G. A. Heterogeneity of aging: individual risk factors, mechanisms, patient priorities and outcomes. *J. Am. Geriatr. Soc.* **69**, 610–612 (2021).
10. Nguyen, Q. D. et al. Health heterogeneity in older adults: exploration in the Canadian longitudinal study on aging. *J. Am. Geriatr. Soc.* **69**, 678–687 (2021).
11. Ibanez, A. & Zimmer, E. R. Time to synergize mental health with brain health. *Nat. Mental Health* **1**, 441–443 (2023).
12. Pang, J. C. et al. Geometric constraints on human brain function. *Nature* **618**, 566–574 (2023).
13. Antal, B. et al. Type 2 diabetes mellitus accelerates brain aging and cognitive decline: complementary findings from UK Biobank and meta-analyses. *eLife* **11**, e73138 (2022).
14. Brouwer, R. M. et al. Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nat. Neurosci.* **25**, 421–432 (2022).
15. Douaud, G. et al. SARS-CoV-2 is associated with changes in brain structure in UK Biobank. *Nature* **604**, 697–707 (2022).
16. Cole, J. H. & Franke, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci.* **40**, 681–690 (2017).
17. Boyle, R. et al. Brain-predicted age difference score is related to specific cognitive functions: a multi-site replication analysis. *Brain Imaging Behav.* **15**, 327–345 (2021).
18. Cole, J. H. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle and cognitive factors. *Neurobiol. Aging* **92**, 34–42 (2020).
19. de Lange, A. G. & Cole, J. H. Commentary: correction procedures in brain-age prediction. *Neuroimage Clin.* **26**, 102229 (2020).
20. Constantinides, C. et al. Brain ageing in schizophrenia: evidence from 26 international cohorts via the ENIGMA Schizophrenia Consortium. *Mol. Psychiatry* **28**, 1201–1209 (2023).
21. Han, L. K. M. et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol. Psychiatry* **26**, 5124–5139 (2021).
22. Guggenmos, M. et al. Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Transl. Psychiatry* **7**, 1279 (2017).
23. Zorina-Lichtenwalter, K. et al. Genetic predictors of human chronic pain conditions. *Neuroscience.* **338**, 36–62 (2016).
24. Yimer, B. B. et al. Heterogeneity in the association between weather and pain severity among patients with chronic pain: a Bayesian multilevel regression analysis. *Pain Rep.* **7**, e963 (2022).
25. Peek, A. L. et al. Increased GABA+ in people with migraine, headache and pain conditions—a potential marker of pain. *J. Pain* **22**, 1631–1645 (2021).
26. López-Otín, C. et al. Hallmarks of aging: an expanding universe. *Cell* **186**, 243–278 (2023).
27. Huang, S. W. et al. Osteoarthritis increases the risk of dementia: a nationwide cohort study in Taiwan. *Sci. Rep.* **5**, 10145 (2015).
28. Innes, K. E. & Sambamoorthi, U. The association of perceived memory loss with osteoarthritis and related joint pain in a large Appalachian population. *Pain Med.* **19**, 1340–1356 (2018).
29. Du, J. et al. Association of APOE-ε4, osteoarthritis, β-amyloid and tau accumulation in primary motor and somatosensory regions in Alzheimer disease. *Neurology* **101**, e40–e49 (2023).
30. Hung, P. S. et al. Differential expression of a brain aging biomarker across discrete chronic pain disorders. *Pain* **163**, 1468–1478 (2022).
31. Yu, G. Z. et al. Accelerated brain aging in chronic low back pain. *Brain Res.* **1755**, 147263 (2021).

32. Kaufmann, T. et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617–1623 (2019).

33. Tachmazidou, I. et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet.* **51**, 230–236 (2019).

34. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

35. Ferreira, L. K. & Busatto, G. F. Resting-state functional connectivity in normal brain aging. *Neurosci. Biobehav. Rev.* **37**, 384–400 (2013).

36. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018).

37. Grubman, A. et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097 (2019).

38. Anderson, K. M. et al. Convergent molecular, cellular and cortical neuroimaging signatures of major depressive disorder. *Proc. Natl Acad. Sci. USA* **117**, 25138–25149 (2020).

39. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).

40. Beck, D. et al. Cardiometabolic risk factors associated with brain age and accelerate brain ageing. *Hum. Brain Mapp.* **43**, 700–720 (2022).

41. Mishra, S., Beheshti, I. & Khanna, P. A review of neuroimaging-driven brain age estimation for identification of brain disorders and health conditions. *IEEE Rev. Biomed. Eng.* **16**, 371–385 (2023).

42. de Lange, A. G. et al. Mind the gap: performance metric evaluation in brain-age prediction. *Hum. Brain Mapp.* **43**, 3113–3129 (2022).

43. Lisman, J. et al. Viewpoints: how the hippocampus contributes to memory, navigation and cognition. *Nat. Neurosci.* **20**, 1434–1447 (2017).

44. Caillaud, M. et al. Evidence of a relation between hippocampal volume, white matter hyperintensities, and cognition in subjective cognitive decline and mild cognitive impairment. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **75**, 1382–1392 (2020).

45. Vetere, G. et al. An inhibitory hippocampal–thalamic pathway modulates remote memory retrieval. *Nat. Neurosci.* **24**, 685–693 (2021).

46. Roh, J. H. et al. Volume reduction in subcortical regions according to severity of Alzheimer's disease. *J. Neurol.* **258**, 1013–1020 (2011).

47. Jahn, H. Memory loss in Alzheimer's disease. *Dialogues Clin. Neurosci.* **15**, 445–454 (2013).

48. Bangen, K. J. et al. Baseline white matter hyperintensities and hippocampal volume are associated with conversion from normal cognition to mild cognitive impairment in the Framingham offspring study. *Alzheimer Dis. Assoc. Disord.* **32**, 50–56 (2018).

49. Jokinen, H. et al. Global burden of small vessel disease-related brain changes on MRI predicts cognitive and functional decline. *Stroke* **51**, 170–178 (2020).

50. Chen, H. H. et al. The diagnostic accuracy of the Ascertain Dementia 8 questionnaire for detecting cognitive impairment in primary care in the community, clinics and hospitals: a systematic review and meta-analysis. *Fam. Pract.* **35**, 239–246 (2018).

51. Nebert, D. W. & Liu, Z. SLC39A8 gene encoding a metal ion transporter: discovery and bench to bedside. *Hum. Genomics* **13**, 51 (2019).

52. Hermann, E. R. et al. Brain magnetic resonance imaging phenome-wide association study with metal transporter gene SLC39A8. *Front. Genet.* **12**, 647946 (2021).

53. Grabert, K. et al. Microglial brain region-dependent diversity and selective regional sensitivities to aging. *Nat. Neurosci.* **19**, 504–516 (2016).

54. Patir, A. et al. A core transcriptional signature of human microglia: derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. *Glia* **67**, 1240–1253 (2019).

55. Fjell, A. M. & Walhovd, K. B. Structural brain changes in aging: courses, causes and cognitive consequences. *Rev. Neurosci.* **21**, 187–221 (2010).

56. Elahi-Mahani, A. et al. Glial cells modulate hippocampal synaptic plasticity in morphine dependent rats. *Brain Res. Bull.* **140**, 97–106 (2018).

57. Takeda, A. Zinc signaling in the hippocampus and its relation to pathogenesis of depression. *J. Trace Elem. Med. Biol.* **26**, 80–84 (2012).

58. Leng, F. & Edison, P. Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat. Rev. Neurol.* **17**, 157–172 (2021).

59. Stephenson, J. et al. Inflammation in CNS neurodegenerative diseases. *Immunology* **154**, 204–219 (2018).

60. Kapoor, M. et al. Role of proinflammatory cytokines in the pathophysiology of osteoarthritis. *Nat. Rev. Rheumatol.* **7**, 33–42 (2011).

61. Papenberg, G. et al. Physical activity and inflammation: effects on gray-matter volume and cognitive decline in aging. *Hum. Brain Mapp.* **37**, 3462–3473 (2016).

62. Marsland, A. L. et al. Interleukin-6 covaries inversely with hippocampal grey matter volume in middle-aged adults. *Biol. Psychiatry* **64**, 484–490 (2008).

63. Maggio, N. & Vlachos, A. Tumor necrosis factor (TNF) modulates synaptic plasticity in a concentration-dependent manner through intracellular calcium stores. *J. Mol. Med.* **96**, 1039–1047 (2018).

64. Paolicelli, R. C. et al. Synaptic pruning by microglia is necessary for normal brain development. *Science* **333**, 1456–1458 (2011).

65. Weerasekera, A. et al. Thalamic neurometabolite alterations in patients with knee osteoarthritis before and after total knee replacement. *Pain* **162**, 2014–2023 (2021).

66. Nakatomi, Y. et al. Neuroinflammation in patients with chronic fatigue syndrome/myalgic encephalomyelitis: an ¹¹C-(R)-PK11195 PET study. *J. Nucl. Med.* **55**, 945–950 (2014).

67. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).

68. Flint, C. et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology* **46**, 1510–1517 (2021).

69. Goyal, M. S. et al. Persistent metabolic youth in the aging female brain. *Proc. Natl Acad. Sci. USA* **116**, 3251–3255 (2019).

70. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).

71. Franke, K. et al. In vivo biomarkers of structural and functional brain development and aging in humans. *Neurosci. Biobehav. Rev.* **117**, 142–164 (2020).

72. Yu, J., Li, J. & Huang, X. The Beijing version of the Montreal Cognitive Assessment as a brief screening tool for mild cognitive impairment: a community-based study. *BMC Psychiatry* **12**, 156 (2012).

73. Gong, Y. & Wang, D. J. *Handbook of Wechsler Memory Scale-Revised* (Hunan Medical College, 1989).

74. Galvin, J. E. et al. The AD8: a brief informant interview to detect dementia. *Neurology* **65**, 559–564 (2005).

75. Broce, I. J. et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. *Acta Neuropathol.* **137**, 209–226 (2019).

76. Elvsåshagen, T. et al. The genetic architecture of the human thalamus and its overlap with ten common brain disorders. *Nat. Commun.* **12**, 2909 (2021).

77. Karadag, N. et al. Identification of novel genomic risk loci shared between common epilepsies and psychiatric disorders. *Brain* **146**, 3392–3403 (2023).

78. Andreassen, O. A. et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am. J. Hum. Genet.* **92**, 197–209 (2013).

79. Andreassen, O. A., Thompson, W. K. & Dale, A. M. Boosting the power of schizophrenia genetics by leveraging new statistical tools. *Schizophr. Bull.* **40**, 13–17 (2014).

80. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).

81. Markello, R. D. et al. Standardizing workflows in imaging transcriptomics with the abagen toolbox. *eLife* **10**, e72129 (2021).

82. Burt, J. B. et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).

83. Huntenburg, J. M., Bazin, P. L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).

84. Burt, J. B. et al. Generative modeling of brain maps with spatial autocorrelation. *Neuroimage* **220**, 117038 (2020).

85. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).

86. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nat. Commun.* **11**, 3358 (2020).

87. Korotkevich, G. et al. Fast gene set enrichment analysis. *BioRxiv* https://doi.org/10.1101/060012 (2016).

88. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

89. Martins, D. et al. Imaging transcriptomics: convergent cellular, transcriptomic and molecular neuroimaging signatures in the healthy adult human brain. *Cell Rep.* **37**, 110173 (2021).

Acknowledgements

Author contributions

Y.T. and J.L. conceived and designed the study. L.Z., W.Z., J.C., T.G. and J.F. analyzed and interpreted the data. J.L. was involved in data collection. L.Z. and Y.T. wrote the initial manuscript. L.Z., Y.T. and T.G. were involved in writing, reviewing and editing the manuscript. Y.T. supervised the project. All authors edited and contributed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-024-00223-3.

Correspondence and requests for materials should be addressed to Yiheng Tu.

Peer review information *Nature Mental Health* thanks Agustin Ibanez and the other, anonymous, reviewers for their contribution to the peer review of this work.

[1]CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China. [2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China. [3]School of Traditional Chinese Medicine, Capital Medical University, Beijing, China. [4]National-Local Joint Engineering Research Center of Rehabilitation Medicine Technology, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, China. [5]School of Data Science, The Chinese University of Hong Kong, Shenzhen, China. [6]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [8]Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [9]These authors contributed equally: Lei Zhao, Jiao Liu. ✉e-mail: yihengtu@gmail.com

nature portfolio

Corresponding author(s):   Yiheng Tu

Last updated by author(s):   Feb 17, 2024

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	The MRI data of Local community was collected on a 3.0T GE 750 scanner (General Electric, Milwaukee, WI, USA) with a 8 channel receive head coil. The MRI data of UK Biobank is directly available from UK Biobank.
Data analysis	MRI data was analyzed using SPM (http://www.fil.ion.ucl.ac.uk/spm) and CAT12 (http://www.neuro.uni-jena.de/cat); Elastic net regression and SVR was performed using Matlab (ww2.mathworks.cn/products/statistics.html); Genetic pleiotropy analysis was performed using pleiofdr toolbox (https://github.com/precimed/pleiofdr); The data of AHBA was preprocessed using abagen toolbox (version 0.1.3,https://github.com/rmarkello/abagen); The spatial autocorrelation in gene transcriptional analysis was corrected using BrainSMASH package (version 0.11.0, https://brainsmash.readthedocs.io/en/latest/index.html); The cell enrichment analyses were conducted using clusterProfiler package (version 3.10.1,https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html); The probe-to-gene reannotation is performed based on Re-annotator toolkit (version 1.0.0,http://sourceforge.net/projects/reannotator); The Custom code can be acquired from github web (https://github.com/tulab-brain/BrainAgeCP).

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

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

This project corresponds to UK Biobank application ID 71901. Data from UK Biobank are publicly available from https://biobank.ndph.ox.ac.uk/upon third-party authorization: https://www.ukbiobank.ac.uk/register-apply. The GWAS summary statistics of chronic pain can be downloaded from https://gwas.mrcieu.ac.uk/datasets with the catagory codes as ukb-b-133, ukb-b-8463, and ukb-b-16118. The GWAS summary statistics of knee osteoarthritis (KOA) can be downloaded from https://www.ebi.ac.uk/gwas/publications/30664745. Human brain gene expression data in AHBA can be downloaded from http://human.brain-map.org/static/download. Individual data used to estimate brain age in Dataset 2 can be accessed via ScienceDB (https://doi.org/10.57760/sciencedb.psych.00120).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	All participants' sex were determined based on self-reporting. The self-reporting sex (Field ID 31) In the UK Biobank was obtained via questionaire.
Reporting on race, ethnicity, or other socially relevant groupings	No socially relevant categorization variables were included in the present study.
Population characteristics	UK Biobank: total number: 9152. age: 62.96±7.70; 45-81. sex: 4259 male and 4893 female. group: 6725 healthy subjects; 982 chronic knee pain, 591 chronic back pain, 528 chronic neck pain, and 326 chronic hip pain subjects. Local community: total number: 192. age: 59.62±6.84; 40-70. sex: 50 male and 142 female. group: 133 KOA and 59 healthy subjects.
Recruitment	The participants of UK Biobank were recruited through postal invitations. The participants from local community were recruited via advertisements posted on social media like WeChat. The recruitment have been biased towards older as aging is the subject of the present study.
Ethics oversight	This study was approved by the Human Research Ethics Committee at the Institute of Psychology of Chinese Academy of Sciences (ethical approval number is H21030) UKB has been approved to obtain and disseminate data and samples by the North West Multi-Centre Research Ethics Committee (http://www.ukbiobank.ac.uk/ethics), and these ethical regulations cover the work in the present study. The study conducted on local community obtained approval from the Medical Ethics Committee of the Affiliated Rehabilitation Hospital of Fujian University of Traditional Chinese Medicine and the Second affiliated hospital of Fujian University of Traditional Chines Medicine.

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

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	No statistical method was used to predetermine sample sizes. The primary analyses included 9,152 participants directly acquired from a open

Sample size	dataset, UK Biobank, making it statistically powerful enough to detect small effect. Moreover, we replicated the fingings using an independent dataset, enhancing the credibility of the conclusions.
Data exclusions	Exclusion criteria for the healthy control group in UK Biobank: 1) without T1-MRI data; 2) have self-reported fair or poor health ratings; 3) have cancer; 4) have non-cancer illnesses; 5) have any self-reported longstanding disability or infirmity; 6) have any self-reported chronic pain condition.

Exclusion criteria for the chronic pain cohorts in UK Biobank: 1) without T1-MRI data; 2) have been included in the healthy control group; 3) have cancer; 4) have heart attacks, angina or stroke diagnosed by doctor; 5) have any self-reported longstanding disability or infirmity; 6) have dementia, Alzheimer's disease, cognitive impairment, epilepsy, cerebral palsy, neuro injury/trauma, depression, anxiety/panic attacks, nervous breakdown, schizophrenia, self-harm/suicide attempt, mania/bipolar disorder/manic depression, post-traumatic stress disorder, alcohol dependency, or other substance abuse/dependency: 7) have self-reported excellent or good health ratings; 8) have more than one type of chronic pain conditions.

Exclusion criteria for the healthy control group in local community: 1) without T1-MRI data; 2) have any history of stroke, severe cerebrovascular disease, musculoskeletal system disease, or sports injury related contraindications; 3) a score of <24 on the cognitive screening by the Mini-Mental State Exam (MMSE); 4) a score of ≥14 on the Beck Depression Inventory (BDI).

Exclusion criteria for the KOA patients in local community: 1) without T1-MRI data; 2) had history of knee surgery within the past 6 months or intra-articular injection of corticosteroids within the past 3 months; 3) knee pain attributable to rheumatic or other inflammatory disease; 4) have primary or secondary muscle disease; 5) have abnormal mental state, not autonomous, or unable to cooperate with study interventions and staff; 6) MRI contraindications, such as dentures, porcelain teeth, and pacemakers 7) have bleeding disorders; 8) have a score of <24 on the cognitive screening by the Mini-Mental State Exam;

Replication	The findings were biologically replicated one time using an independent dataset, including increased predicted age difference (PAD) in KOA, the contribution of the hippocampus to PAD, and transcriptional associations between the KOA neuroimaging phenotypes and expression of gene SLC39A8.
Randomization	In our statistical analyses, we regressed out covariates including sex, age, and total intracranial volume (TIV). The educational levels, medications, and comorbidities were adjusted in the sensitivity analyses
Blinding	Blinding was not applicable to this study as this is observational.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	*State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	*For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Reporting on sex	*Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No | Yes

☐ | ☐ Public health

☐ | ☐ National security

☐ | ☐ Crops and/or livestock

☐ | ☐ Ecosystems

☐ | ☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No | Yes

☐ | ☐ Demonstrate how to render a vaccine ineffective

☐ | ☐ Confer resistance to therapeutically useful antibiotics or antiviral agents

☐ | ☐ Enhance the virulence of a pathogen or render a nonpathogen virulent

☐ | ☐ Increase transmissibility of a pathogen

☐ | ☐ Alter the host range of a pathogen

☐ | ☐ Enable evasion of diagnostic/detection modalities

☐ | ☐ Enable the weaponization of a biological agent or toxin

☐ | ☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	Structural MRI (T1), Diffution MRI, and resting-state functional MRI.

| Design specifications | UK Biobank designed imaging acquisition protocols including 6 modalities. The collection order is T1-weighted structural imaging, resting-state functional MRI, task functional MRI, T2-weighted FLAIR structural imaging, Diffution MRI, and susceptibility-weighted imaging.

Local community designed imaging acquisition protocols including 2 modalities. The collection order is T1-weighted structural imaging and resting-state functional MRI. |
|---|---|
| Behavioral performance measures | For the KOA patients in local community, cognitive function and memory function were measured by the Montreal Cognitive Function Assessment Scale (MoCA) and Wechsler Memory Scale-Chinese Revision (WMS-CR). The AD8 questionnaire was used to evaluate dementia risk in KOA patients. The knee injury and osteoarthritis outcome scores (KOOS) was used to evaluate pain characteristics. For the KOA individuals with relaxed inclusion criteria in UKB, memory function were measured by a numeric memory test (Data-Field 4282). |

Acquisition

Imaging type(s)	T1-weighted structural imaging, Diffusion imaging, and Functional imaging.
Field strength	3T
Sequence & imaging parameters	For UK Biobank, detailed sequence and imaging parameters can be obtained from an open-source document (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf).

For local community, T1-weighted structural imaging were acquired on a 3.0T GE 750 scanner and 8-channel receive head coil (General Electric, Milwaukee, WI, USA) with the following parameters: slice thickness = 1 mm, flip angle = 15°, field of view (FOV) = 240 mm, and 160 slices in acquisition. |
| Area of acquisition | Whole brain for T1-weighted structural imaging; ROI-based imaging phenotypes for Diffusion and Functional imaging (directly acquired from UK Biobank). |
| Diffusion MRI | ☒ Used   ☐ Not used |

Preprocessing

Preprocessing software	T1-weighted structural imaging data was preprocessed using SPM12 and Computational Anatomy Toolbox (CAT12)；Diffusion and functional imaging data was directly acquired from UK Biobank.
Normalization	Non-linear
Normalization template	MNI152NLin2009cAsym_Geodesic Shooting templates
Noise and artifact removal	As part of the CAT12 default pipeline
Volume censoring	Not applicable

Statistical modeling & inference

Model type and settings	Multivariate analyses were used to estimate brain age; Univariate analyses were used to compare PAD and relate it to behavior and genetic data.
Effect(s) tested	Pearson's correlation coefficient and mean absolute error between brain age and chronological age was used to assess the brain age model; Two-sample t-tests were used to perform between-group inference on PAD, PAD changes during follow-up, and grey matter volume (GMV). Equivalence testing was used to examine the equivalence between the longitudinal change of the predicted age and chronological age. Pearson correlation was used to examine the associations between PAD and GMV, cognitive function and pain characteristics. Spearman correlation was used to examine the associations between PAD and AD8 scores. Fisher's Exact Test was used to examine the associations between diagnosis of dementia and PAD.
Specify type of analysis:	☐ Whole brain ☐ ROI-based ☒ Both
Statistic type for inference

(See Eklund et al. 2016) | Voxel-wise analysis was performed to develop MRI(VBM)-based brain age model; ROI-wise analysis was performed to develop SC- and FC-based brain age model; ROI-wise analysis was performed to correlate MRI phenotypes with gene expression. |
| Correction | False discovery rate (FDR) and Bonferroni |

Models & analysis

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

Functional and/or effective connectivity | Pearson correlation

Graph analysis | *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis | A elastic net regression predictive model was used to estimate brain age. The performance of the model was assessed using Pearson's correlation coefficient and mean absolute error (MAE).