nature neuroscience

BRIEF COMMUNICATION

https://doi.org/10.1038/s41593-019-0471-7

Corrected: Publisher Correction

Common brain disorders are associated with heritable patterns of apparent aging of the brain

Tobias Kaufmann[1*], Dennis van der Meer[1,2], Nhat Trung Doan[1], Emanuel Schwarz[3], Martina J. Lund[1], Ingrid Agartz[1,4,5], Dag Alnæs[1], Deanna M. Barch[6,7,8], Ramona Baur-Streubel[9], Alessandro Bertolino[10,11], Francesco Bettella[1], Mona K. Beyer[12,13], Erlend Bøen[4,14], Stefan Borgwardt[15,16,17], Christine L. Brandt[1], Jan Buitelaar[18,19], Elisabeth G. Celius[12,20], Simon Cervenka[5], Annette Conzelmann[21], Aldo Córdova-Palomera[1], Anders M. Dale[22,23,24,25], Dominique J. F. de Quervain[26,27], Pasquale Di Carlo[11], Srdjan Djurovic[28,29], Erlend S. Dørum[1,30,31], Sarah Eisenacher[3], Torbjørn Elvsåshagen[1,12,20], Thomas Espeseth[30], Helena Fatouros-Bergman[5], Lena Flyckt[5], Barbara Franke[32], Oleksandr Frei[1], Beathe Haatveit[1,30], Asta K. Håberg[33,34], Hanne F. Harbo[12,20], Catharina A. Hartman[35], Dirk Heslenfeld[36,37], Pieter J. Hoekstra[38], Einar A. Høgestøl[12,20], Terry L. Jernigan[39,40,41], Rune Jonassen[42], Erik G. Jönsson[1,5], Karolinska Schizophrenia Project (KaSP)[43], Peter Kirsch[44,45], Iwona Kłoszewska[46], Knut K. Kolskår[1,30,31], Nils Inge Landrø[4,30], Stephanie Le Hellard[29], Klaus-Peter Lesch[47,48,49], Simon Lovestone[50], Arvid Lundervold[51,52], Astri J. Lundervold[53], Luigi A. Maglanoc[1,30], Ulrik F. Malt[12,54], Patrizia Mecocci[55], Ingrid Melle[1], Andreas Meyer-Lindenberg[3], Torgeir Moberget[1], Linn B. Norbom[1,30], Jan Egil Nordvik[56], Lars Nyberg[57], Jaap Oosterlaan[36,58], Marco Papalino[11], Andreas Papassotiropoulos[26,59,60], Paul Pauli[9], Giulio Pergola[11], Karin Persson[61,62], Geneviève Richard[1,30,31], Jaroslav Rokicki[1,30], Anne-Marthe Sanders[1,30,31], Geir Selbæk[12,61,62], Alexey A. Shadrin[1], Olav B. Smeland[1], Hilkka Soininen[63,64], Piotr Sowa[13], Vidar M. Steen[29,65], Magda Tsolaki[66], Kristine M. Ulrichsen[1,30,31], Bruno Vellas[67], Lei Wang[68], Eric Westman[16,69], Georg C. Ziegler[47], Mathias Zink[3,70], Ole A. Andreassen[1] and Lars T. Westlye[1,30*]

Common risk factors for psychiatric and other brain disorders are likely to converge on biological pathways influencing the development and maintenance of brain structure and function across life. Using structural MRI data from 45,615 individuals aged 3–96 years, we demonstrate distinct patterns of apparent brain aging in several brain disorders and reveal genetic pleiotropy between apparent brain aging in healthy individuals and common brain disorders.

Psychiatric disorders and other brain disorders are among the main contributors to morbidity and disability around the world[1]. The disease mechanisms are complex, spanning a wide range of contributing genetic and environmental factors[2]. The inter-individual variability is large, but on a group level, patients with common brain disorders perform worse on cognitive tests, are less likely to excel professionally, and engage in adverse health behaviors more frequently than healthy individuals[3]. It is unclear to what extent these characteristics are a cause, consequence or confounder of disease.

Dynamic processes that influence the rate of brain maturation and change throughout the lifespan have a critical role, as reflected in the wide range of times of disease onset from early childhood to old age[4]. This suggests that the age at which individual trajectories diverge from the norm reflects key characteristics of the underlying pathophysiology. Although autism spectrum disorder (ASD) and attention-deficit hyperactivity disorder (ADHD) emerge in childhood[5], schizophrenia and bipolar spectrum disorders are likely to develop during late childhood and adolescence, before the characteristic outbreak of severe symptoms in early adulthood[6]. Likewise, multiple sclerosis most often manifests in early adulthood, but the disease process probably starts much earlier[7]. First episodes in major depressive disorder (MDD) can appear at any stage from adolescence to old age[5], whereas mild cognitive impairment (MCI) and dementia primarily emerge during senescence[8]. Beyond such differential temporal evolution across the lifespan, age-related deviations from the norm may also differ between disorders in terms of anatomical location, direction, change rate and magnitude, all of which add complexity to the interpretation of observed effects.

Machine learning techniques enable robust estimation of the biological age of the brain using information provided by MRI[9,10], assessing the similarity of a given brain scan with scans of a range of individuals to estimate the age of the tissue from a normative

A full list of affiliations appears at the end of the paper.

lifespan trajectory. Initial evidence suggested that the deviation between brain age and chronological age—termed the brain age gap—is a promising marker of brain health[11], but several issues remain to be addressed. First, although advantageous for narrowing the complexity, reducing a rich set of brain imaging features into a single estimate of brain age inevitably compromises spatial specificity, thereby neglecting disorder-specific patterns. Second, most studies so far have been small-scale, were performed within a limited age range and have focused on a single disorder, which rendered them unable to uncover clinical specificity and lifespan dynamics. Third, the genetic underpinnings of brain age gap are not understood, and it is unknown to what extent they overlap with the genetic architecture of major clinical traits. To address these critical knowledge gaps, large imaging genetics samples covering a range of prevalent brain disorders are necessary.

Here, we used a centralized and harmonized processing protocol including automated surface-based morphometry and subcortical segmentation using Freesurfer on raw structural MRI data from 45,615 individuals aged 3–96 years that passed quality control (Supplementary Fig. 1). The sample included data from healthy controls ($n=39{,}827$, aged 3–95 years) and 5,788 individuals with various brain disorders. We included data from individuals with ASD ($n=925$, 5–64 years), ADHD ($n=725$, 7–62 years), prodromal schizophrenia or at-risk mental state (SZRISK, $n=94$, 16–42 years), schizophrenia ($n=1110$, 18–66 years), a heterogeneous group with mixed diagnoses in the psychosis spectrum (PSYMIX, $n=300$, 18–69 years), bipolar spectrum disorder ($n=459$, 18–66 years), multiple sclerosis ($n=254$, 19–68 years), MDD ($n=208$, 18–71 years), MCI ($n=974$, 38–91 years) and dementia (including Alzheimer's disease, $n=739$, 53–96 years). Supplementary Tables 1–3 provide details on sample characteristics and scanning protocols.

We used machine learning to estimate individual brain age on the basis of structural brain imaging features. First, we grouped all subjects into different samples. For each of the ten clinical groups, we identified a group of healthy individuals of equal size, matched by age, sex and scanning site from a pool of 4,353 healthy controls. All remaining individuals were combined into one independent sample comprising only healthy individuals. This independent sample constituted a training sample, used to train and tune the machine learning models for age prediction ($n=35{,}474$, aged 3–89 years, 18,990 female participants), whereas the ten clinical samples were used as independent test samples. Figure 1a illustrates the respective age distributions per sex and diagnosis.

The large sample size and wide age-span of the training sample allowed the male and female brain age to be modeled separately, thereby accounting for potential sexual dimorphisms in brain structural lifespan trajectories[12]. For each sex, we built a machine learning model based on gradient tree boosting to predict the age of the brain from a set of thickness, area and volume features extracted using a multimodal parcellation of the cerebral cortex as well as a set of cerebellar–subcortical volume features (1,118 features in total, Fig. 1b). Fivefold cross-validations revealed a high degree of correlation between chronological age and predicted brain age ($r=0.93$ for the female model and $r=0.94$ for the male model, Supplementary Fig. 2). Supplementary Figs. 3–6 provide further validation of the prediction approach and Supplementary Table 4 provides details on sex differences in the prediction models. Next, we applied the models to predict brain age for each individual in the ten independent test samples (predicting brain age using the female model in female subjects and the male model in male subjects) and tested for effects of diagnosis on the brain age gap using linear models. We used mega-analysis (that is, across-site analysis) as the main statistical framework and provide results from a meta-analysis framework in the Supplementary Information. We included age, age^2, sex, scanning site and a proxy of image quality (Euler number) in all statistical models testing for group differences and clinical associations. To further minimize confounding effects of data quality, we repeated the main analyses using a more stringent quality control and exclusion procedure.

Figure 2a illustrates that the estimated brain age gap was increased in several brain disorders. The strongest effects were observed in schizophrenia (Cohen's $d=0.51$), multiple sclerosis ($d=0.74$), MCI ($d=0.41$) and dementia ($d=1.03$). PSYMIX ($d=0.21$) and bipolar spectrum disorder ($d=0.29$) showed small effects of increased brain age gap, whereas other groups showed negligible effects ($d<0.2$). The meta-analysis converged on the same findings (Supplementary Fig. 7) and the results were replicated regardless of the quality control exclusion criterion applied (Supplementary Fig. 8). The brain age gap in all clinical groups was positive on average and there were no signs of a negative brain age gap (developmental delay) in children with ASD or ADHD; there was also no significant group-by-age interaction effect (Supplementary Table 5).

We assessed the specificity of the spatial brain age gap patterns across clinical groups. We trained age prediction models using only occipital, frontal, temporal, parietal, cingulate, insula, or cerebellar–subcortical features (Fig. 1b). Cross-validation confirmed the predictive performance of all regional models (Supplementary Fig. 2), which were used to predict regional brain age in the ten independent test sets. Regional brain age gaps largely corresponded to the full brain level, with some notable differential spatial patterns (Fig. 2b). For example, increased cerebellar–subcortical age gap was most prominent in dementia ($d=0.99$) and multiple sclerosis ($d=0.81$) but was not present in schizophrenia ($d=0.16$). The largest effect in schizophrenia was observed in the frontal lobe ($d=0.70$). A brain age gap in the temporal lobe was observed in MDD ($d=0.24$), whereas there was no evidence ($d<0.2$) for a brain age gap in ASD, ADHD or SZRISK in any of the regions. To explore regional differences in brain age patterns, we tested for group-by-region interactions on each pairwise combination of clinical groups and pairwise combination of regional brain age gaps (1,260 tests). Figure 2c illustrates the significant effect sizes, indicating that the rate at which different regions age in relation to each other often showed opposite patterns in disorders typically considered neurodevelopmental (for example, schizophrenia) and neurodegenerative (for example, multiple sclerosis or dementia).

With converging evidence demonstrating the largest brain age gaps in schizophrenia, multiple sclerosis, MCI and dementia, we explored the functional relevance of the regional brain age gaps for these groups by testing for associations with clinical and cognitive data. Clinical data available from individuals with schizophrenia included symptom ($n=389$) and function ($n=269$) scores of the Global Assessment of Functioning scale (GAF) as well as positive ($n=646$) and negative ($n=626$) scores of the Positive and Negative Syndrome Scale (PANSS). For multiple sclerosis, we assessed associations with scores from the Expanded Disability Status Scale (EDSS, $n=195$). In the dementia spectrum, we assessed associations with Mini Mental State Examination scores (MMSE, $n=907$ MCI, $n=686$ dementia). Figure 2d depicts association strengths accounting for age, age^2, sex, scanning site and Euler number and Supplementary Fig. 11 provides corresponding scatter plots. In schizophrenia, larger brain age gaps were associated with lower functioning (for example, full brain age gap with GAF symptom ($r=-0.15$, $P=.003$) and insula brain age gap with GAF function ($r=-0.22$, $P=3\times10^{-4}$)) and with more negative symptoms (for example, temporal brain age gap with PANSS negative ($r=0.13$, $P=.001$)). In multiple sclerosis, larger full brain age gap was associated with a higher degree of disability ($r=0.23$, $P=.001$). Finally, lower cognitive functioning was associated with larger brain age gaps in MCI or dementia, with strongest effects for full brain ($r=-0.30$, $P=7\times10^{-33}$) and cerebellar–subcortical ($r=-0.29$, $P=2\times10^{-30}$) brain age gaps.

Given the substantial genetic contributions to most brain disorders, our results incite the question to what degree brain age

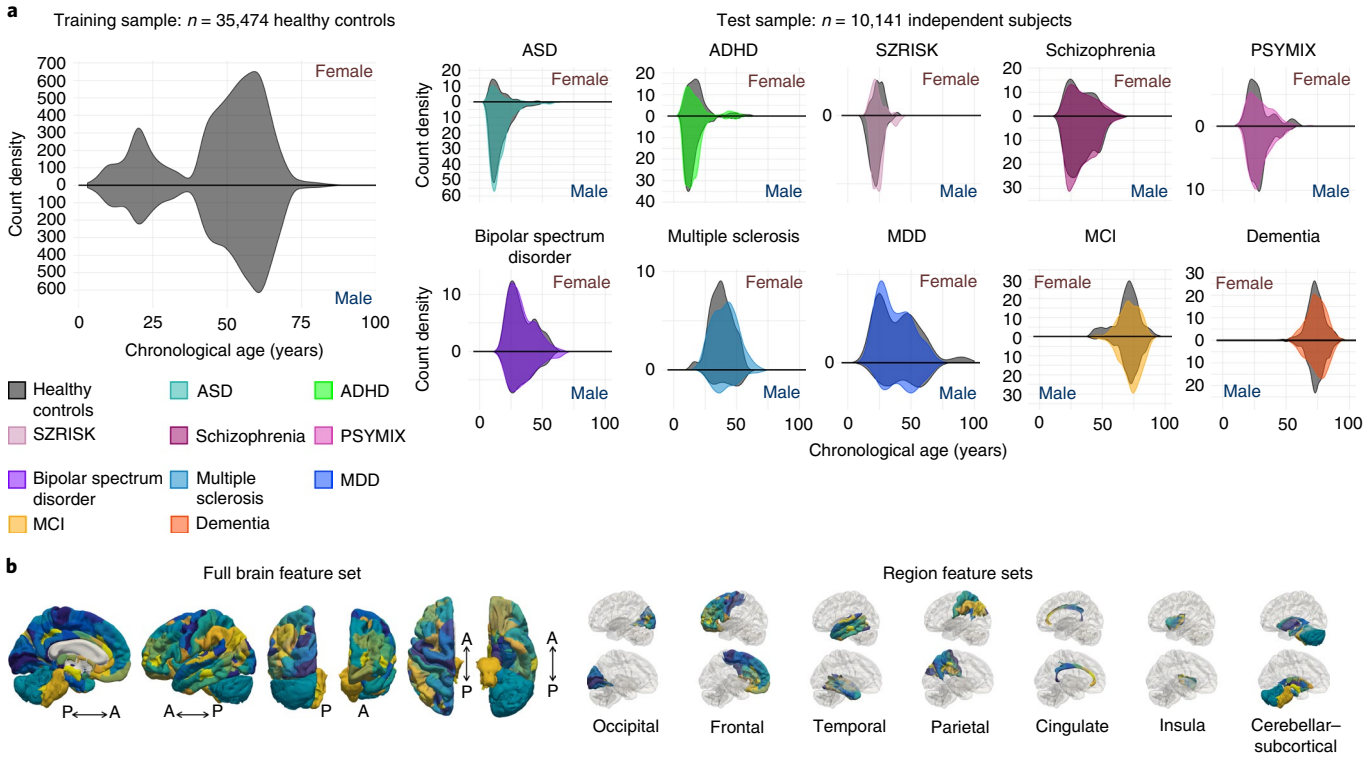

Fig. 1 | Sample distributions and imaging features used for brain age prediction. a, Age distributions of the training (left) and the ten test (right) samples per sex and diagnosis. The gray shading behind each clinical group reflect its age-, sex- and site-matched control group. **b**, Cortical features from the Human Connectome Project (HCP) atlas as well as cerebellar–subcortical features used for brain age prediction. A, anterior; P, posterior. Colors were assigned randomly to each feature. All features were used in the full brain feature set (left), whereas only those from specific regions (occipital, frontal, temporal, parietal, cingulate, insula and cerebellar–subcortical) were included in the regional feature set (right). For illustrative purposes, the left hemisphere is shown.

patterns are genetically influenced and whether the implicated polymorphisms overlap with the polygenic architectures of the disorders. We used single nucleotide polymorphism (SNP) data from the 20,170 adult healthy individuals with European ancestry available in the UK Biobank. We estimated full and regional brain age for these individuals using fivefold cross-validation in models trained on all healthy controls ($n = 39{,}827$ aged 3–95 years; 20,868 female participants, models trained per sex).

First, we performed one genome-wide association study (GWAS) per brain age gap using PLINK, including the first ten population components from multidimensional scaling, age, age^2, sex, scanning site and Euler number as covariates. Next, we assessed heritability using linkage disquilibrium (LD) score regression on the resulting summary statistics. In line with earlier results from twin studies[13], our SNP-based analysis revealed significant heritability (Fig. 3a), with common SNPs explaining 24% of the variance in brain age gap across all individuals (full brain, $h^2_{SNP} = 0.24$, s.e. = 0.03) and 17–23% of the variance in regional brain age gaps (all s.e. < 0.03).

Next, we assessed the overlap between the genetic underpinnings of brain age gap and common brain disorders. We gathered GWAS summary statistics for ASD, ADHD, schizophrenia, bipolar spectrum disorder, multiple sclerosis, MDD and Alzheimer's disease (see Methods). First, using LD score regression, we assessed the genetic correlation between these summary statistics and those from brain age gaps. Correlations were overall weak (Supplementary Fig. 12), with only one surviving false-discovery rate (FDR) correction for the number of tests (cingulate brain age gap with ADHD). Lack of genetic correlation does not preclude genetic dependence because traits may have mixed effect directions across shared genetic variants[14]. Therefore, we next used conjunctional FDR analyses to identify SNPs that are significantly associated with both brain age gap and disorders. We found significant independent loci showing pleiotropy between brain age gaps and all included disorders (Fig. 3b). Most loci were identified for schizophrenia (two occipital, four frontal, three temporal, six parietal, five cingulate, five insula and two cerebellar–subcortical; 161 SNPs in total). Further, five independent loci for ASD (76 SNPs), six for ADHD (80 SNPs), ten for bipolar spectrum disorder (94 SNPs), five for multiple sclerosis (22 SNPs), one for MDD (14 SNPs) and six for Alzheimer's disease (15 SNPs) (Fig. 3c). Supplementary Table 6 provides details. An intronic variant in protein coding gene *SATB2* at chromosome 2q33.1 was most frequently associated with brain age gaps and schizophrenia. A missense variant in protein coding gene *SLC39A8* was associated with cerebellar–subcortical brain age gap and schizophrenia and showed the strongest effect in all tested associations ($P = 9 \times 10^{-8}$).

Taken together, our results provide strong evidence that several common brain disorders are associated with an apparent aging of the brain, with effects observed at the full brain or regional level in schizophrenia, PSYMIX, bipolar spectrum disorder, multiple sclerosis, MDD, MCI and dementia; but not in ASD, ADHD or SZRISK. Importantly, our approach revealed differential neuroanatomical distribution of brain age gaps between several disorders. Associations with clinical and cognitive patient data supported the functional relevance of the brain age gaps and genetic analyses in healthy individuals provided evidence that the brain age gaps are heritable, with overlapping genes between brain age gaps in healthy adults and common brain disorders.

Our approach of estimating regional brain age was useful to reveal differential spatial patterns between disorders. Although the implicated regions in the spatial brain age profiles of the disorders

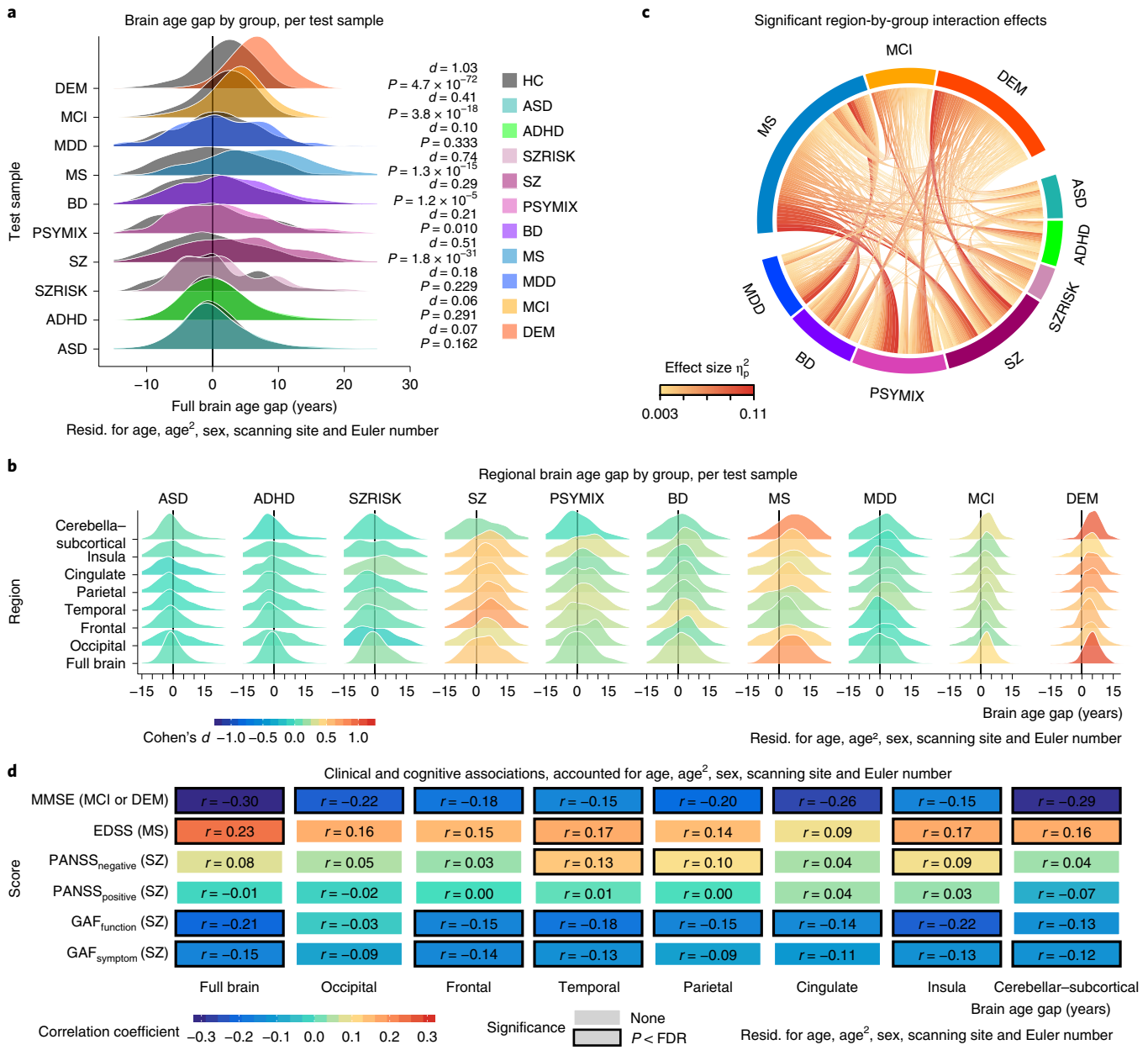

Fig. 2 | Apparent brain aging is common in several brain disorders and is sensitive to clinical and cognitive measures. a, The gap between chronological age and brain age was increased in several disorders. The gray shading behind each clinical group reflect its age-, sex- and site-matched controls. The test samples comprised $n = 925$ ASD, $n = 925$ healthy controls; $n = 725$ ADHD, $n = 725$ healthy controls; $n = 94$ SZRISK, $n = 94$ healthy controls; $n = 1{,}110$ schizophrenia, $n = 1{,}110$ healthy controls; $n = 300$ PSYMIX, $n = 300$ healthy controls; $n = 459$ bipolar spectrum disorder, $n = 459$ healthy controls; $n = 254$ multiple sclerosis, $n = 254$ healthy controls; $n = 208$ MDD, $n = 208$ healthy controls; $n = 974$ MCI, $n = 974$ healthy controls; $n = 739$ dementia, $n = 739$ healthy controls (in total, $n = 10{,}141$ independent subjects). Cohen's d effect sizes (pooled s.d. units) and two-sided P values are provided. **b**, Several disorders showed specific patterns in regional brain age gaps. Colors indicate Cohen's d effect sizes for group comparisons. Sample size as specified in **a**. A corresponding correlation matrix of the effect sizes is depicted in Supplementary Fig. 9. **c**, Effect sizes of significant region-by-group interactions from repeated-measures ANOVA run for each combination of regions and groups (1,260 tests in total). Sample size as specified in **a** but excluding healthy controls ($n = 5{,}788$ independent subjects). Only significant ($P <$ FDR, Benjamini–Hochberg) effects are shown. Supplementary Fig. 10 depicts effect sizes for all 1,260 tests. **d**, Correlation coefficients for linear associations between brain age gaps and cognitive and clinical scores. Sample size comprised $n = 389$ schizophrenia for $GAF_{symptom}$, $n = 269$ schizophrenia for $GAF_{function}$, $n = 646$ schizophrenia for $PANSS_{positive}$, $n = 626$ schizophrenia for $PANSS_{negative}$, $n = 195$ multiple sclerosis for EDSS, $n = 907$ MCI and $n = 686$ dementia for MMSE. Associations were computed using linear models accounting for age, age², sex, scanning site and Euler number, and the resulting t-statistics were transformed to r. Significant ($P <$ FDR, Benjamini–Hochberg, two-sided) associations are marked with a black box. Corresponding scatter plots are depicted in Supplementary Fig. 11. BD, bipolar spectrum disorder; DEM, dementia; MS, multiple sclerosis; SZ, schizophrenia; HC, healthy controls, resid., residualized.

largely corresponded with previously reported structural abnormalities (for example, frontal in schizophrenia[15] and substantial subcortical volume loss in Alzheimer's disease[16]), our regional brain age approach preserved the well-established benefit of down-sampling a large number of brain imaging features into a condensed and interpretable score without a total loss of spatial

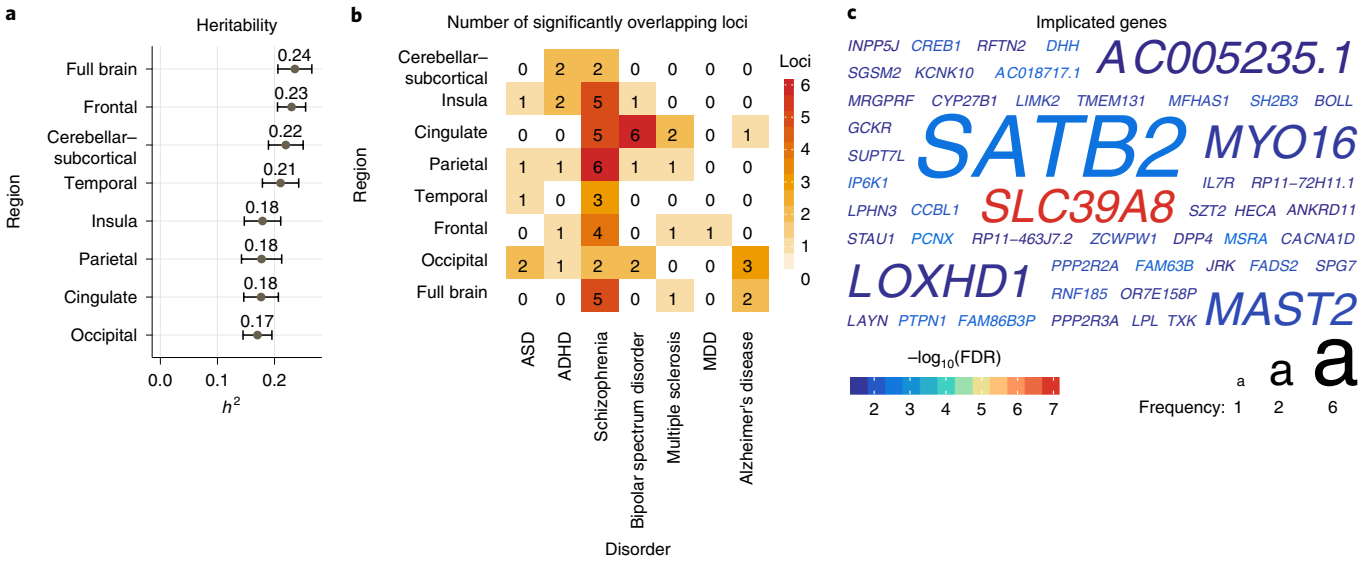

Fig. 3 | The brain age gaps are heritable, and the genetic underpinnings overlap with those observed for several disorders. Genetic analyses were performed using data from 20,170 healthy adult individuals with European ancestry. **a**, Heritability (h^2) estimated using LD score regression. Error bars, s.e. **b**, Significantly ($P <$ FDR) overlapping loci between brain age gaps and disorders, identified using conjunctional FDR. **c**, Corresponding to **b**, the overlapping genes across all disorders, with color indicating significance and size indicating frequency of detection.

sensitivity. As such, the analysis revealed substantial differences in spatial aging profiles between disorders typically regarded as neurodegenerative (multiple sclerosis, MCI and dementia) and neurodevelopmental (in particular, schizophrenia and PSYMIX). For example, although these disorders were all associated with an increased brain age gap on the full brain level, regional analysis revealed interactions between the frontal brain age patterns observed in schizophrenia and the cerebellar–subcortical patterns observed in multiple sclerosis and dementia, supporting spatial differences in apparent brain age. Moreover, significant associations with clinical and cognitive data, in particular with scores of the GAF and PANSS in schizophrenia, with the EDSS in multiple sclerosis and with MMSE in the dementia spectrum, demonstrated the functional relevance of the brain age gap beyond group differences. By gauging the dynamic associations between changes in brain age and clinical and cognitive function, future longitudinal studies may prove instrumental to dissect the large individual differences among patients with brain disorders, even within the same diagnostic category[17]. Furthermore, incorporating additional imaging modalities, voxel-level data or different segmentations at various levels of resolution will allow for estimation of tissue-specific brain age gaps or different regional gaps in future studies. Such approaches will also be useful to further investigate the apparent lack of brain age gap differences in ASD and ADHD. In contrast to research from other imaging phenotypes[18,19], we did not observe case-control differences in brain age gaps for ASD or ADHD, nor group-by-age interactions (developmental delays might be reflected in a negative brain age gap in children). Brain age gaps based on different imaging modalities may capture different aspects of pathophysiology and will therefore make an important contribution in future research.

Conceptually, brain age gaps reflect a prediction error from a machine learning model and can therefore be attributed to both noise (lack of model accuracy, insufficient data quality) and physiology (deviations from normal aging trajectories). The large training sample and accurate model performance, replication of results using different data quality criteria, as well as our approach of comparing brain age gaps of cases to a group of age-, sex- and scanner-matched controls allowed us to reduce the impact of noise and to attribute variation in brain age gaps as probably related to biologically relevant differences. The physiological underpinnings of the brain age gaps are likely to be diverse, much like the polygenic nature of brain disorders and their profoundly heterogeneous symptomatology. They may reflect differences in disease severity, or effects of comorbid disorders, substance use or other adverse lifestyle factors. Genetic analysis offers one way of exploring factors that influence phenotypic variation towards an improved understanding of the multi-faceted sources of lifespan trajectories in the brain. Here, we have provided evidence that full and regional brain age gaps represent genetically influenced traits, and have illustrated that the genetic variants associated with brain age gaps in healthy individuals partly overlap with those observed in ASD, ADHD, schizophrenia, bipolar spectrum disorder, multiple sclerosis, MDD and Alzheimer's disease. In line with accumulating evidence that common brain disorders are highly polygenic and partly overlapping[20], these results suggest shared molecular genetic mechanisms between brain age gaps and brain disorders. Statistical associations do not necessarily signify causation, and functional interpretations of the identified genes should be made with caution. Larger imaging genetics samples, in particular those including individuals with common brain disorders, may in the future allow investigation of the specificity of the implicated genes, and integrating a wider span of imaging modalities may increase both sensitivity and specificity.

In conclusion, we have established that the brain age gap is increased in several common brain disorders, is sensitive to clinical and cognitive phenotypes and is genetically influenced. Our results emphasize the potential of advanced lifespan modeling in the clinical neurosciences, highlighting the benefit of big data resources that cover a wide age-span and conditions. Delineating dynamic lifespan trajectories within and across individuals will be essential to disentangle the pathophysiological complexity of brain disorders.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0471-7.

Received: 28 August 2018; Accepted: 22 July 2019;
Published online: 24 September 2019

References

1. WHO. World Health Statistics 2016: monitoring health for the SDGs. https://www.who.int/gho/publications/world_health_statistics/2016/en/ (2016).
2. Insel, T. R. & Cuthbert, B. N. Brain disorders? Precisely. *Science* **348**, 499–500 (2015).
3. Prince, M. et al. No health without mental health. *Lancet* **370**, 859–877 (2007).
4. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet* **16**, 441–458 (2015).
5. Marin, O. Developmental timing and critical windows for the treatment of psychiatric disorders. *Nat. Med* **22**, 1229–1238 (2016).
6. Insel, T. R. Rethinking schizophrenia. *Nature* **468**, 187–193 (2010).
7. Aubert-Broche, B. et al. Onset of multiple sclerosis before adulthood leads to failure of age-expected brain growth. *Neurology* **83**, 2140–2146 (2014).
8. Masters, C. L. et al. Alzheimer's disease. *Nat. Rev. Dis. Primers* **1**, 15056 (2015).
9. Dosenbach, N. U. et al. Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361 (2010).
10. Franke, K., Ziegler, G., Klöppel, S. & Gaser, C. & Alzheimera's Disease Neuroimaging Initiative. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *Neuroimage* **50**, 883–892 (2010).
11. Cole, J. H. & Franke, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci.* **40**, 681–690 (2017).
12. Ritchie, S. J. et al. Sex differences in the adult human brain: evidence from 5216 UK biobank participants. *Cereb. Cortex* **28**, 2959–2975 (2018).
13. Cole, J. H. et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage* **163**, 115–124 (2017).
14. Bansal, V. et al. Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nat. Commun.* **9**, 3078 (2018).
15. Ellison-Wright, I. & Bullmore, E. Anatomy of bipolar disorder and schizophrenia: a meta-analysis. *Schizophr. Res.* **117**, 1–12 (2010).
16. Jernigan, T. L., Salmon, D. P., Butters, N. & Hesselink, J. R. Cerebral structure on MRI, part II: specific changes in Alzheimer's and Huntington's diseases. *Biol. psychiatry* **29**, 68–81 (1991).
17. Wolfers, T. et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *Jama Psychiat* **75**, 1146–1155 (2018).
18. Ecker, C., Bookheimer, S. Y. & Murphy, D. G. Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *Lancet Neurol.* **14**, 1121–1134 (2015).
19. Faraone, S. V. et al. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primers* **1**, 15020 (2015).
20. Andreassen, O. A. et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Mol. psychiatry* **20**, 207 (2015).

Acknowledgements

The author list between I.A. and M.Z. is in alphabetic order. The authors were funded by the Research Council of Norway (276082 LifespanHealth (T.K.), 213837 (O.A.A), 223273 NORMENT (O.A.A.), 204966 (L.T.W.), 229129 (O.A.A.), 249795 (L.T.W.), 273345 (L.T.W.) and 283798 SYNSCHIZ (O.A.A.)), the South-Eastern Norway Regional Health Authority (2013-123 (O.A.A.), 2014-097 (L.T.W.), 2015-073 (L.T.W.) and 2016083 (L.T.W.)), Stiftelsen Kristian Gerhard Jebsen (SKGJ-MED-008), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Starting Grant, Grant agreement No. 802998 BRAINMINT (L.T.W.)), NVIDIA Corporation GPU Grant (T.K.), and the European Commission 7th Framework Programme (602450, IMAGEMEND (A.M.-L.)). The data used in this study were gathered from various sources. A detailed overview of the included cohorts and acknowledgement of their respective funding sources and cohort-specific details is provided in Supplementary Table 1. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu), the AddNeuroMed consortium and the Pediatric Imaging, Neurocognition and Genetics (PING) study database (www.chd.ucsd.edu/research/ping-study.html, now shared through the NIMH Data Archive (NDA)). The investigators within the ADNI and PING studies contributed to the design and implementation of ADNI and PING or provided data but did not participate in the analysis or writing of this report. This publication is solely the responsibility of the authors and does not necessarily represent the views of the National Institutes of Health or PING investigators. Complete listings of participating sites and study investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf and http://pingstudy.ucsd.edu/investigators.html. The AddNeuroMed consortium was led by S.L., B.V., P.M., M.T., I.K. and H.S.

Author contributions

T.K. and L.T.W. conceived the study, T.K., N.T.D. and L.T.W. pre-processed all data in Freesurfer, N.T.D., M.J.L., C.L.B, L.B.N., L.T.W. and T.K. performed quality control of the data, T.K. performed the analysis with contributions from L.T.W. and D.v.d.M., and T.K., L.T.W., N.T.D., D.v.d.M. and O.A.A. contributed to interpretation of the results. All remaining authors were involved in data collection at various sites as well as cohort-specific tasks. T.K. and L.T.W. wrote the first draft of the paper and all authors contributed to and approved the final manuscript.

Competing interests

Some authors received educational speaker's honoraria from Lundbeck (O.A.A., A.B., T.E., M.Z., N.I.L.), Sunovion (O.A.A.), Shire (B.F.), Medice (B.F.), Otsuka (A.B., M.Z.), Janssen (A.B.), Roche (M.Z.), Ferrer (M.Z.), Trommsdorff (M.Z.) and Servier (M.Z.), all unrelated to this work. A.B. is a stockholder of Hoffmann-La Roche and has received consultant fees from Biogen Idec. S.L. is currently an employee of Janssen-Cilag, but contribution to this work was completed prior to this employment. E.A.H., E.G.C., M.K.B., P.S., and H.F.H. have received travel support, honoraria for advice and/or lecturing from Almirall (E.G.C.), Biogen Idec (E.G.C., H.F.H., M.K.B.), Sanofi-Genzyme (E.G.C., H.F.H., E.A.H.), Merck (E.G.C., H.F.H., E.A.H., P. S.), Novartis (E.G.C., H.F.H., M.K.B.), Roche (E.G.C., H.F.H.), Sanofi-Aventis (E.G.C., H.F.H.) and Teva (E.G.C., H.F.H.). E.G.C. anf H.F.H. have received unrestricted research grants from Novartis (E.G.C., H.F.H.), Biogen Idec (E.G.C.) and Sanofi-Genzyme (E.G.C.). G.P. has been the academic supervisor of a Roche collaboration grant (years 2015-2016) that funds his salary. None of the mentioned external parties had any role in the analysis, writing or decision to publish this work. All other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0471-7.

Correspondence and requests for materials should be addressed to T.K. or L.T.W.

Peer review information *Nature Neuroscience* thanks Janine Bijsterbosch, Gagan Wig, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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

[1]NORMENT, Division of Mental Health and Addiction Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [2]School of Mental Health and Neuroscience Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands. [3]Department of Psychiatry and Psychotherapy Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [4]Department of Psychiatry Diakonhjemmet Hospital, Oslo, Norway. [5]Centre for Psychiatry Research, Department of Clinical Neuroscience Karolinska Institutet & Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden. [6]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, USA. [7]Department of Psychiatry Washington, University in St. Louis, St. Louis, USA. [8]Department of Radiology Washington, University in St. Louis, St. Louis, USA. [9]Department of Psychology I, University of Würzburg, Würzburg, Germany. [10]Institute of Psychiatry Bari University Hospital, Bari, Italy. [11]Department of Basic Medical Science, Neuroscience and Sense Organs University of Bari, Bari, Italy. [12]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [13]Division of Radiology and Nuclear Medicine, Section of Neuroradiology Oslo University Hospital, Oslo, Norway. [14]Psychosomatic and CL Psychiatry, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [15]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [16]Department of Psychiatry, Psychosomatics and Psychotherapy University of Lübeck, Lübeck, Germany. [17]Institute of Psychiatry King's College, London, UK. [18]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour Radboud University Medical Center, Nijmegen, The Netherlands. [19]Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands. [20]Department of Neurology, Oslo University Hospital, Oslo, Norway. [21]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy University of Tübingen, Tübingen, Germany. [22]Center for Multimodal Imaging and Genetics, University of California at San Diego, La Jolla, CA, USA. [23]Department of Radiology, University of California, San Diego, La Jolla, CA, USA. [24]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. [25]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. [26]Division of Cognitive Neuroscience, University of Basel, Basel, Switzerland. [27]Transfaculty Research Platform Molecular and Cognitive Neurosciences University of Basel, Basel, Switzerland. [28]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [29]NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway. [30]Department of Psychology, University of Oslo, Oslo, Norway. [31]Sunnaas Rehabilitation Hospital HT, Nesodden, Norway. [32]Departments of Human Genetics and Psychiatry, Donders Institute for Brain, Cognition and Behaviour Radboud University Medical Center, Nijmegen, The Netherlands. [33]Department of Neuromedicine and Movement Science Norwegian, University of Science and Technology, Trondheim, Norway. [34]Department of Radiology and Nuclear Medicine St. Olavs Hospital, Trondheim, Norway. [35]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [36]Clinical Neuropsychology section Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [37]Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [38]Department of Child and Adolescent Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [39]Center for Human Development, University of California, San Diego, USA. [40]Department of Cognitive Science, University of California, San Diego, USA. [41]Departments of Psychiatry and Radiology, University of California, San Diego, USA. [42]Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway. [43]A list of authors and affiliations appears at the end of the paper. [44]Department of Clinical Psychology Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [45]Bernstein Center for Computational Neuroscience Heidelberg/Mannheim, Mannheim, Germany. [46]Department of Old Age Psychiatry and Psychotic Disorders Medical University of Lodz, Lodz, Poland. [47]Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Würzburg, Germany. [48]Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine Sechenov First Moscow State Medical University, Moscow, Russia. [49]Department of Neuroscience, School for Mental Health and Neuroscience (MHeNS) Maastricht University, Maastricht, The Netherlands. [50]Department of Psychiatry, Warneford Hospital University of Oxford, Oxford, UK. [51]Department of Biomedicine, University of Bergen, Bergen, Norway. [52]Mohn Medical Imaging and Visualization Centre, Department of Radiology, Haukeland University Hospital, Bergen, Norway. [53]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway. [54]Department of Research and Education, Oslo University Hospital, Oslo, Norway. [55]Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy. [56]CatoSenteret Rehabilitation Center Son, Oslo, Norway. [57]Departments of Radiation Sciences and Integrative Medical Biology, Umeå Center for Functional Brain Imaging Umeå University, Umeå, Sweden. [58]Emma Children's Hospital, Amsterdam UMC University of Amsterdam and Vrije Universiteit Amsterdam, Emma Neuroscience Group, Department of Pediatrics, Amsterdam Reproduction & Development, Amsterdam, The Netherlands. [59]Division of Molecular Neuroscience University of Basel, Basel, Switzerland. [60]Life Sciences Training Facility, Department Biozentrum University of Basel, Basel, Switzerland. [61]Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway. [62]Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway. [63]Department of Neurology, Institute of Clinical Medicine University of Eastern Finland, Kuopio, Finland. [64]Neurocenter, Neurology Kuopio University Hospital, Kuopio, Finland. [65]Dr. E. Martens Research Group for Biological Psychiatry, Department of Medical Genetics Haukeland University Hospital, Bergen, Norway. [66]1st Department of Neurology Aristotle University of Thessaloniki, Thessaloniki, Greece. [67]UMR Inserm 1027, CHU Toulouse, UPS, Toulouse, France. [68]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [69]Department of Neurobiology Care Sciences and Society, Karolinska Institute, Stockholm, Sweden. [70]District hospital Ansbach, Ansbach, Germany. *e-mail: bias.kaufmann@medisin.uio.no; l.t.westlye@psykologi.uio.no

Karolinska Schizophrenia Project (KaSP)

Lars Farde[5], Lena Flyckt[5], Göran Engberg[71], Sophie Erhardt[71], Helena Fatouros-Bergman[5], Simon Cervenka[5], Lilly Schwieler[71], Fredrik Piehl[72], Ingrid Agartz[1,4,5], Karin Collste[5], Pauliina Victorsson[5], Anna Malmqvist[71], Mikael Hedberg[71] and Funda Orhan[71]

[71]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden. [72]Neuroimmunology Unit, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.

Methods

Samples. We have included data collected through collaborations, data sharing platforms and consortia, as well as available in-house cohorts. No statistical methods were used to pre-determine sample sizes. We included as much data as we could gather (brain scans from 45,615 individuals) and the sample size of individual clinical groups is therefore based on data availability. Supplementary Tables 1–3 provide detailed information on the individual cohorts. All included cohorts have been reported on previously, and we refer to a list of publications that can be consulted for a more detailed overview of cohort characteristics. Data collection in each cohort was performed with participants' written informed consent and with approval by the respective local Institutional Review Boards.

Image pre-processing and quality control. Raw T1 data for all study participants were stored and analysed locally at the University of Oslo (Norway), following a harmonized analysis protocol applied to individual subject data (Supplementary Fig. 1). We performed automated surface-based morphometry and subcortical segmentation using Freesurfer 5.3 (ref. [21]). We deployed an automated quality control protocol executed within each of the contributing cohorts that excluded potential outliers based on the Euler number[22] of the respective Freesurfer segmentations. The Euler number captures the topological complexity of the uncorrected Freesurfer surfaces and is thus a proxy of data quality[22]. In brief, for each scanning site we regressed age, age^2 and sex from the Euler number of the left and right hemispheres and identified scans that deceeded 3 standard deviations on either of the residualized Euler numbers. Supplementary Fig. 13 provides a validation of the approach against manual quality control. Data from a total of 977 individuals was excluded in this step, yielding 45,615 individuals for the main analysis. To further minimize the confounding effects of data quality[23], we performed supplementary analyses using a subset of data, for which a more stringent threshold was used for exclusion (1 s.d. on Euler numbers). Thus, supplemental analysis provides a confirmation with those individuals excluded ($n = 40{,}301$ remaining).

Brain age prediction. We used a recent multimodal cortical parcellation scheme[24] to extract cortical thickness, area and volume for 180 regions of interest (ROIs) per hemisphere. In addition, we extracted the classic set of cerebellar–subcortical and cortical summary statistics[21]. This yielded a total set of 1,118 structural brain imaging features (360 cortical thickness, 360 cortical area, 360 cortical volume and 38 cerebellar–subcortical and cortical summary statistics).

We used machine learning on this feature set to predict the age of each individual's brain. First, we split the available data into a training sample and ten independent test samples (Fig. 1a). The test samples in total comprised 5,788 individuals with brain disorders and 4,353 healthy controls. For each of the ten clinical groups, we selected a set of healthy controls from the pool of 4,353 individuals, matched for age, sex and scanning site using propensity score matching[25]. Thus, data from some healthy individuals acted as control data in several test samples, yet each test sample had the same number of patients and controls and all subjects in the test samples were independent of the subjects in the training sample. The remaining datasets ($45{,}615 - (5{,}788 + 4{,}353) = 35{,}474$) went into the training set. For each sex, we trained machine learning models based on gradient tree boosting[26] using the *xgboost* package in R[27], which was chosen owing to its resource efficiency and demonstrated superior performance in previous machine learning competitions[26], to predict the age of the brain using data available in the training set. First, model parameters were tuned using a fivefold cross-validation of the training data. This step identified the optimal number of model training iterations by assessing the prediction error for 1,500 rounds and implementing early stopping if the performance did not improve for 20 rounds. Based on previous experience, the learning rate was pre-set to $\eta = 0.01$ and all other parameters were set to default[27] for linear xgboost tree models. After determining the optimal number of training iterations, the full set of training data was used to train the final models with the adjusted nrounds parameter. These models were used to predict brain age in the test samples, and the brain age gap (deviation between brain and chronological age) was computed. In line with a recent recommendation[28], all statistical analyses on the brain age gap accounted for age, age^2, sex, scanning site and Euler number. In addition, to assess overall model performance, a fivefold cross-validation was performed within the training set, with each fold implementing the above described training procedure and testing on the hold-out part of the training set. Brain age predictions on the level of individual brain regions followed the same procedures as those described for the full brain level, except that the feature set was reduced to cover only those features that overlapped more than 50% with a given region. Regions were defined following the Freesurfer lobesStrict segmentation as occipital, frontal, temporal, parietal, cingulate and insula. In addition, given the limited number of cerebellar features available in the Freesurfer summary statistics, cerebellar and subcortical features were grouped into a cerebellar–subcortical region (Fig. 1b). For additional validation, we compared our xgboost approach against two other approaches (Supplementary Fig. 3). One approach implemented a different machine learning algorithm on the same set of features (slm from the care package[29]), whereas the other approach made use of a fully independent processing pipeline, feature set and algorithm (https://github.com/james-cole/brainageR[13,30]). Furthermore, we assessed the impact of sample size on model performance by creating random subsets of data with sample sizes of 100, 500, 1,000, 2,000, 5,000, 10,000 and 20,000 individuals (40 random subsets per sample size). For each subset and sample size we assessed model performance using cross-validation (Supplementary Fig. 5).

The genetic analysis was performed using UK Biobank data, which was part of the training set in the main analysis. We thus trained different brain age models for the genetic analysis. We selected all healthy controls and estimated their brain age using a fivefold cross-validation approach, similar to the one performed when validating the performance of the training set. The resulting unbiased estimates of brain age gaps for all UK Biobank individuals with genetic data available went into the genome-wide association analysis, LD score regression and conjunctional FDR.

Main statistical analysis framework. We performed both mega- (across cohorts) and meta- (within cohort) analyses. To estimate group effects on a given measure in a mega-analysis framework, we computed the effect of diagnosis in relation to the healthy controls for each of the ten test samples in a linear model accounting for age, age^2, sex, scanning site and Euler number. Cohen's d effect sizes were estimated based on contrast t-statistics[31] following equation (1):

$$d = \frac{t(n_1 + n_2)}{\sqrt{n_1 n_2}\sqrt{df}} \quad (1)$$

$$v = \left(\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2 - 2)}\right)\left(\frac{n_1 + n_2}{n_1 + n_2 - 2}\right) \quad (2)$$

For the meta-analysis, similar models were computed within cohorts. In addition to estimating Cohen's d using equation (1), we estimated the variance of d following equation (2).

Cumulative effects across cohorts were then estimated using a variance-weighted random-effects model as implemented in the metafor package in R[32].

Data distributions were assumed to be normal, but this was not formally tested. Data collection and analysis were not performed blind to the conditions of the experiments.

Assessment of regional specificity. In Supplementary Fig. 9, we performed clustering of effect sizes from Fig. 2b using heatmap.2 from the gplots package[33] in R. A Spearman correlation matrix was computed based on the case-control effect sizes obtained from each test sample and region and hierarchical clustering was performed using the default settings. To further explore regional specificity, we performed an analysis that involved only the clinical groups. We regressed age, age^2, sex, scanning site and Euler number from the brain age gaps in each test sample. Next, we joined data from each pair of clinical groups and each pair of regions for repeated measures ANOVA and estimated the effect sizes of region-by-group interactions (1,260 ANOVAs in total). The significant interaction effects were visualized in Fig. 2c using the circlize package[34] in R 3.6.

Genetic analyses. We restricted all genetic analyses to individuals from the UK Biobank with European ancestry, as determined by the UK Biobank study team[35]. We applied standard quality control procedures to the UK Biobank v3 imputed genetic data. In brief, we removed SNPs with an imputation quality score below 0.5, with a minor allele frequency less than 0.05, missing in more than 5% of individuals, and failing the Hardy Weinberg equilibrium tests at $P < 1 \times 10^{-6}$, yielding SNP data from 20,170 adult healthy individuals. We performed a genome-wide association analysis using PLINK v1.9 (ref. [36]), accounting the analysis for ten genetic principal components, age, age^2, sex, scanning site and Euler number. We used LD score regression[37] to estimate narrow sense heritability.

Furthermore, we used cross-trait LD score regression[37,38] to calculate genetic correlations, and conjunctional FDR analyses[39,40] to assess genetic overlap between two complex traits. We gathered genome-wide association analysis summary statistics for ASD[41], ADHD[42], schizophrenia[43], bipolar spectrum disorder[44], multiple sclerosis[45], major depression[46], and Alzheimer's disease[47] and assessed genetic overlap with brain age gap genetics. The major histocompatibility complex (MHC) region was excluded from all analyses. Conjunctional FDR was run for each pair of full brain or regional brain age gap and group, using a conjunctional FDR threshold of 0.05. SNPs were annotated using the Ensembl Variant Effect Predictor[48].

Cognitive and clinical associations. Cognitive and clinical associations were tested in subsets based on data availability and were performed in clinical groups only (excluding controls), as described in the main text. Using linear models accounting for age, age^2, sex, scanning site and Euler number we associated brain age gaps with scores of the GAF scale[49], the PANSS[50], the EDSS[51] and the MMSE scores[52]. The t-statistics of the linear models were transformed to r; therefore, the correlation coefficients depicted in Fig. 2d essentially reflect a partial correlation between full brain or regional brain age gaps and clinical or cognitive scores, controlling for confounding effects of age, sex, site and image quality.

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

Data availability

The raw data incorporated in this work were gathered from various resources. Material requests will need to be placed with individual principal investigators. A detailed overview of the included cohorts is provided in Supplementary Table 1. GWAS summary statistics for the brain age gaps as well as the models needed to predict brain age in independent cohorts are available at github.com/tobias-kaufmann.

Code availability

Code needed to run brain age prediction models is available at github.com/tobias-kaufmann (see Data availability). Additional R statistics[53] code is available from the authors upon request.

References

21. Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
22. Rosen, A. F. G. et al. Quantitative assessment of structural image quality. *Neuroimage* **169**, 407–418 (2018).
23. Smith, S. M. & Nichols, T. E. Statistical challenges in "Big Data" human neuroimaging. *Neuron* **97**, 263–268 (2018).
24. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
25. Ho, D., Imai, K., King, G. & Stuart, E. A. MatchIt: nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* **42**, 1–28 (2011).
26. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
27. Chen, T., et al. Xgboost: extreme gradient boosting. R package v0.4-2 https://cran.r-project.org/web/packages/xgboost/ (2015).
28. Le, T. T. et al. A nonlinear simulation framework supports adjusting for age when analyzing BrainAGE. *Front. Aging Neurosci.* **10**, 317 (2018).
29. Zuber, V. & Strimmer, K. Care. R package v 1.1.10. https://cran.r-project.org/web/packages/care/care.pdf (2017).
30. Cole, J. H. et al. Brain age predicts mortality. *Mol. psychiatry* **23**, 1385–1392 (2018).
31. Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* **82**, 591–605 (2007).
32. Viechtbauer, W. Conducting meta-analysis in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
33. Warnes, G. R. et al. R Package gplots: various R programming tools for plotting data. https://cran.r-project.org/web/packages/gplots/gplots.pdf (2016).
34. Gu, Z. R. Package circlize: circular visualization. https://cran.r-project.org/web/packages/circlize/circlize.pdf (2017).
35. Bycroft, C. et al. TheUK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
36. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
37. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
38. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
39. Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J.-B. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**, 653–660 (2005).
40. 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).
41. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
42. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genet.* **51**, 63–75 (2018).
43. Schizophrenia Working Group of the PGC. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).
44. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
45. Patsopoulos, N. et al. The multiple sclerosis genomic map: role of peripheral immune cells and resident microglia in susceptibility. Preprint at *biorXiv* https://www.biorxiv.org/content/10.1101/143933v1 (2017).
46. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
47. Lambert, J.-C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452 (2013).
48. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
49. Pedersen, G. & Karterud, S. The symptom and function dimensions of the Global Assessment of Functioning (GAF) scale. *Compr. Psychiatry* **53**, 292–298 (2012).
50. Kay, S. R., Fiszbein, A. & Opfer, L. A. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* **13**, 261 (1987).
51. Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* **33**, 1444–1444 (1983).
52. Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
53. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, Vienna, Austria, 2013).

nature research

Corresponding author(s): Tobias Kaufmann and Lars T. Westlye

Last updated by author(s): Jul 20, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	This is an analysis of previously collected magnetic resonnance imaging and genetics data. Cohort-specific details on data collection are provided in Suppl. Tables 1-3 and the references cited therein.
Data analysis	GWAS summary statistics for the brain age gaps as well as the models needed to predict brain age in independent cohorts will be made available at github.com/tobias-kaufmann upon acceptance. Software used to analyse the data: - Freesurfer 5.3 - Custom scripts in R 3.6, using packages xgboost 0.82, ggplot2 3.1, metafor 2.1, care 1.1, gplots 3.0, circlize 0.4, psych 1.8, ggridges 0.5 , MatchIt 3.0 - BrainageR for model comparison (https://github.com/james-cole/brainageR) - PLINK 1.9 - LD Score regression (https://github.com/bulik/ldsc) - Ensembl Variant Effect Predictor (https://www.ensembl.org/info/docs/tools/vep/index.html) - Matlab 2018 with conjunctional FDR 1.4

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data incorporated in this work were gathered from various resources (see acknowledgements). Material requests will need to be placed with individual PIs. Corresponding authors Tobias Kaufmann (tobias.kaufmann@medisin.uio.no) and Lars T. Westlye (l.t.westlye@psykologi.uio.no) will provide additional detail upon correspondence.
Accession codes are provided in Supplementary Table 1.

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	We have included data collected through collaborations, data sharing platforms, consortia as well as available in-house cohorts. No statistical methods were used to pre-determine sample sizes. We included as much data as we could gather (brain scans from N=45,615 individuals) and sample size of individual clinical groups is thus based on data availability.
Data exclusions	Raw T1 data for all study participants were stored and analysed locally at University of Oslo, following a harmonized analysis protocol applied to each individual subject data (Suppl. Fig. 1). We performed automated surface-based morphometry and subcortical segmentation using Freesurfer 5.3. We deployed an automated quality control protocol executed within each of the contributing cohorts that excluded potential outliers based on the Euler number of the respective Freesurfer segmentations. Euler number captures the topological complexity of the uncorrected Freesurfer surfaces and thus comprises a proxy of data quality. In brief, for each scanning site we regressed age, age^2 and sex from the Euler number of the left and right hemispheres and identified scans that deceeded 3 standard deviations (SD) on either of the residualized Euler numbers. Suppl. Fig. 13 provides a validation of the approach against manual quality control. Data from a total of 977 individuals was excluded in this step, yielding 45,615 subjects for the main analysis. To further minimize confounding effects of data quality, we performed supplementary analyses using a subset of data, where a more stringent threshold was used for exclusion (1 SD on Euler numbers). Thus, supplemental analysis provides a sanity check with those subjects excluded (sample size: n = 40,301).
Replication	We split data into a training and test set. We validated the machine learning models using 5-fold cross validation within the training set. After verification of prediction accuracy, we applied the models to the independent test sets. We provide results from mega- and meta-analysis. Furthermore, to ensure that our results would replicate after more stringent outlier exclusion, we performed additional supplemental analysis as described above under "Data exclusions" and replicated the results in the respective subset of data with highest quality.
Randomization	We grouped all subjects into different samples. For each of the ten clinical groups, we identified a group of healthy individuals of equal size, matched on age, sex and scanning site from a pool of 4353 healthy control subjects. All remaining individuals were joined into one independent sample comprising healthy individuals only. The latter constituted a training sample, used to train and tune the machine learning models for age prediction (n = 35,474 aged 3-89 years; 18,990 females), whereas the ten clinical samples were used as independent test samples.
Blinding	The brain age prediction models were tuned within the training set. Thus, the results obtained in clinical groups are drawn from predictions in independent test sets.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study included data from healthy controls (HC; n = 39,827; 3-95 years), as well as from 5788 individuals with diverse brain disorders with typical onset age distributed across the lifespan. We included data from individuals with ASD (n = 925; 5-64 years) and ADHD (n = 725; 7-62 years), individuals with prodromal SZ or at risk mental state (SZRISK; n = 94; 16-42 years), individuals with SZ (n = 1110; 18-66 years), a heterogeneous group with mixed diagnoses in the psychosis spectrum (PSYMIX; n = 300; 18-69 years), individuals with BD (n = 459; 18-66 years), MS (n = 254; 19-68 years), MDD (n = 208; 18-71 years), MCI (n = 974; 38-91 years), and DEM (including Alzheimer's disease; n = 739; 53-96 years). Supplementary Tables 1-3 provide details on the samples' characteristics and scanning protocols.
Recruitment	Cohort-specific details on recruitment are provided in the referenced publications in Supplementary Table 1.
Ethics oversight	This is a re-analysis of previously published data from studies that have each received ethical approval. More information for each individual study is available in the referenced publications in Supplementary Table 1.

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

Magnetic resonance imaging

Experimental design

Design type	Anatomical scan
Design specifications	Information available in Suppl. Table 3 and the therein cites references.
Behavioral performance measures	No task was performed (anatomical scan)

Acquisition

Imaging type(s)	structural, T1-weighted
Field strength	1.5T or 3T, see Suppl. Table 3
Sequence & imaging parameters	Suppl. Table 3
Area of acquisition	whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	We employed a centralized and harmonized processing protocol including automated surface-based morphometry and subcortical segmentation using Freesurfer 5.3 (recon-all)
Normalization	We used standard procedures as implemented in Freesurfer recon-all.
Normalization template	fsaverage
Noise and artifact removal	Standard pipelines for anatomical data were applied (Freesurfer recon-all). Euler number was calculated as a proxy of image quality and data from individuals with insufficient image quality were excluded.
Volume censoring	No volume censoring was performed (anatomical scan)

Statistical modeling & inference

Model type and settings	We used machine learning to predict brain age in an independent test set. Group statistics and association analyses

Model type and settings

were performed on the resulting brain age gaps. We controlled all associations and group differences for age, age², sex, scanning site and a proxy of image quality (Euler number).

Effect(s) tested

We used linear models to assess the effect of group on brain age gap within each test sample, accounting for age, age², sex, scanning site and Euler number. Given the relationship between P-values and sample size, we used Cohen's d effect sizes as the main statistical outcome, but also provide two-sided P-values alongside. Statistical analysis is based on Nakagawa and Cuthill (2007) and described in detail in the online methods section.
Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc 82, 591-605, doi:10.1111/j.1469-185X.2007.00027.x (2007).

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

Anatomical location(s) Glasser et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature

Statistic type for inference (See Eklund et al. 2016)

Linear models on brain age gaps.

Correction

FDR

Models & analysis

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

Multivariate modeling and predictive analysis

We utilized a recent cortical parcellation scheme (Glasser et al, reference above) to extract cortical thickness, area and volume for 180 regions of interest (ROI) per hemisphere. In addition, we extracted the classic set of cerebellar/subcortical and cortical summary statistics. This yielded a total set of 1118 structural brain imaging features (360/360/360/38 for cortical thickness/area/volume as well as cerebellar/subcortical and cortical summary statistics, respectively).
We used machine learning on this feature set to predict the age of each individual's brain. First, we split the available data into a training sample and ten independent test samples (Fig. 1a). The test samples in total comprised 5788 individuals with brain disorders and 4353 healthy controls. For each of the ten clinical groups, we selected a set of healthy controls from the pool of 4353 individuals, matched for age, sex and scanning site using propensity score matching. Thus, data from some healthy individuals acted as control data in several test samples, yet each test sample had the same number of patients and controls and all subjects in the test samples were independent of the subjects in the training sample. The remaining datasets (45,615 – (5788+4353) = 35,474) went into the training set. For each sex, we trained machine learning models based on gradient tree boosting utilizing the xgboost package in R, chosen due to its resource efficiency and demonstrated superior performance in previous machine learning competitions, to predict the age of the brain using data available in the training set. First, model parameters were tuned using a 5-fold cross-validation of the training data. This step identified the optimal number of model training iterations by assessing the prediction error for 1500 rounds and implementing an early stopping if the performance did not improve for 20 rounds. Based on previous experience, the learning rate was pre-set to eta=0.01 and all other parameters were set to default for linear xgboost tree models. After determining the optimal number of training iterations, the full set of training data was used to train the final models with the adjusted nrounds parameter. These models were used to predict brain age in the test samples, and the brain age gap (deviation between brain and chronological age) was computed. In line with a recent recommendation, all statistical analyses on the brain age gap accounted for age, age², sex and scanning site. In addition, to assess overall model performance, prediction models were cross-validated within the training set using a 5-fold cross validation, each fold implementing the above described training procedure and testing on the hold-out part of the training set. Brain age predictions on the level of individual brain regions followed the same procedures as those described for the full brain level, except that the feature set was reduced to cover only those features that overlapped more than 50% with a given lobe. Regions were defined following the Freesurfer lobesStrict segmentation as occipital, frontal, temporal, parietal, cingulate and insula. In addition, given the limited number of cerebellar features available in the Freesurfer summary statistics, cerebellar and subcortical features were grouped into a cerebellar/subcortical region (Fig. 1b). For additional validation, we compared our xgboost approach against two other approaches (Suppl. Fig. 3). One approach implemented a different machine learning algorithm on the same set of features (slm from the care package), whereas the other approach made use of a fully independent processing pipeline, feature set and algorithm (github.com/james-cole/brainageR). Furthermore, we assessed the impact of sample size on model performance by creating random subsets of data with sample sizes of 100, 500, 1000, 2000, 5000, 10,000, and 20,000 individuals (40 random subsets per sample size). For each subset and sample size we assessed model performance using cross-validation [See supplement for references to the methods described above]