Deciphering brain heterogeneity in neuropsychiatric disorders with artificial neural networks

Dissertation for the degree of Philosophiae Doctor (Ph.D)

Esten Høyland Leonardsen
Department of Psychology
Faculty of Social Science

UNIVERSITY OF OSLO
2023

Series of dissertations submitted to the
Faculty of Social Sciences, University of Oslo
No. 1008

ISSN 1504-3991

Cover: UiO.
Print production: Graphic center, University of Oslo.

Acknowledgements

The work performed in this thesis would not have been possible without the support and guidance of those around me. I feel lucky to have been able to work on such an interesting and important topic surrounded by inspiring, knowledgeable, and motivated people. The last four years have been a great experience, and I am grateful to everyone who has made it possible.

First and foremost, I would like to thank my supervisors Yunpeng Wang, Lars T. Westlye, and Thomas Wolfers. Your feedback has been invaluable, our discussions enlightening, and I have learned a lot from all of you. Yunpeng, thank you for guiding me through the process, allowing me to make my own decisions, and supporting me when they have been suboptimal. I am thankful for being given this opportunity and proud of what we have achieved. Thank you, Lars, for sharing your insights through inspiring discussions, and encouraging me to go in my own direction. Thank you, Thomas, for always being accessible when I needed help, and for the great experiences we've had together.

I want to acknowledge UiO:LifeScience for funding this project and applaud their efforts to promote interdisciplinary research. Being put in a context with experts from different fields than my own incurred a steep learning curve and some tough discussions, but also opportunities to work on new problems with new methods that would otherwise have been inconceivable.

I want to thank Kristine B. Walhovd and Anders M. Fjell and everyone else at the Center for Lifespan Changes in Brain and Cognition, for welcoming me into the group and providing an encouraging and motivating work environment throughout these years. I have enjoyed our discussions and learnt a lot from our interactions.

This work was made possible through a variety of data-collection efforts. I want to thank all the participants for volunteering their time to participate in these studies and the researchers who performed the work and made the data accessible.

Finally, I want to thank my family and friends for their support and encouragement. Thank you for allowing me to be a nerd when I have needed to, and providing opportunities to do something else when necessary. In particular, thank you to Elise for ensuring a functional family life and household during periods of late nights and weekends spent working, and enriching my life in between them.

Contents

1 General Summary **3**

2 List of papers **5**

3 Abbreviations **6**

4 Introduction **7**

4.1 The neuroscientific perspective on neuropsychiatric disorders . . 7

4.1.1 Brain imaging in neuropsychiatric disorders 10

4.1.2 Shortcomings of the case-control paradigm in the face of heterogeneity and comorbidity 12

4.1.3 From description to prediction: New venues for modelling the relationship between brain and behaviour 13

4.2 Representation learning and image recognition 18

4.2.1 Deep learning in neuroimaging 20

4.2.2 Brain age and proxy measures 24

4.2.3 Understanding deep neural networks with explainable artificial intelligence 27

4.3 Validating imaging discoveries with genetic analyses 32

4.3.1 Elucidating mechanistic and causal relationships via genetic associations 34

4.4 Summary 35

5 Research objectives **36**

5.1 Paper I 36

5.2 Paper II 36

5.3 Paper III 37

6 Methodology **38**

6.1 Data 38

6.1.1 Imaging data 40

6.1.2 Genetic data 42

6.1.3 Phenotypic data 42

6.2 Modelling 43

6.2.1 Convolutional neural network architectures 43

6.2.2 Model training and comparison 45
6.2.3 Model evaluation 45
6.2.4 Operationalizing brain age 46
6.2.5 Explaining model predictions 47
6.3 Statistical analyses 48
6.3.1 Case-control analyses 48
6.3.2 Genetic analyses 50
6.3.3 Validating relevance maps 51
6.3.4 Additional analyses 52
6.4 Ethical considerations 53

7 Summary of papers **61**
7.1 Paper I: Deep neural networks learn general and clinically relevant representations of the ageing brain 61
7.2 Paper II: Genetic architecture of brain age and its causal relations with brain and mental disorders 62
7.3 Paper III: Characterizing personalized neuropathology in dementia and mild cognitive impairment with explainable artificial intelligence 63

8 Discussion **64**
8.1 Deep learning-derived brain age as a robust marker of generalized brain health 64
8.2 Elucidating brain age via genetic associations 66
8.3 Characterizing heterogeneity with explainable artificial intelligence 67
8.4 The current state of deep learning in neuroimaging 70
8.5 Limitations and methodological considerations 73
8.5.1 Intricacies of clinical validations 73
8.5.2 The limits imposed by existing samples 75
8.5.3 The validity of cross-sectional inferences 76
8.5.4 Architectural choices 77

9 Concluding remarks **79**

References **80**

Papers I-III **123**

1 General Summary

Neuropsychiatric disorders are emerging as leading causes of disability and mortality on a global scale (World Health Organization, 2021, 2022). Many of these are clinical syndromes characterized by a variety of symptoms, including cognitive impairments and behavioural dysfunctions. However, these symptoms are dimensional and often overlap between disorders, complicating the effort to precisely recognize and delineate them. This has ignited a search for quantitative biological measures that can be used to accurately describe aberrations in individual patients, both to elucidate the true nature of these conditions and facilitate precise and personalized clinical decision-making. One potential venue to realize this goal is through Magnetic Resonance Imaging (MRI), where high-resolution images are acquired non-invasively to capture biological aspects of the brain. However, analyses of MRI scans from neuropsychiatric patients have so far been unable to reveal canonical patterns of aberrations distinctive of the different disorders, and discerning them remains a challenge. One possible explanation for this inability is the limited expressive power of the statistical models employed to detect patterns of deviations in the imaging data.

Over the last decade, the renaissance of artificial intelligence and deep learning has provided new opportunities to discover and exploit complex patterns in image data for predictive tasks (LeCun et al., 2015). This has been enabled by artificial neural networks, advanced statistical learning models that learn how to combine facets of data to form new, composite, representations, revealing new perspectives of its content. These technological advances present a pivotal opportunity for detecting subtle and intricate neuroanatomical aberrations in MRI data that were previously inaccessible due to their complexity. This could expand our knowledge of the biological deviations associated with neuropsychiatric disorders and provide accurate tools to support personalized diagnostics and treatments. However, it requires overcoming two major challenges innate to deep learning models: their incomprehensibility, making them hard to trust and learn from, and their demand for large datasets.

In the current thesis, we aimed to explore the potential of using deep neural networks to learn complex representations from neuroimaging data to characterize biological aberrations in patients afflicted by neuropsychiatric disorders. To this end, we trained convolutional neural networks (CNNs) on structural MRI data from large, heterogeneous datasets to predict various tasks. Next,

we used information derived from these models to evaluate their potential for supporting clinical decision-making and to elucidate the association between the representations they have learned and various other measures.

In paper I we trained CNNs to predict apparent brain age, a promising imaging biomarker for depicting generalized brain health, highlighting the capacity of these models to learn robust representations that generalize to unseen data. Next, we demonstrated that the brain age predicted by the model was associated with a range of other phenotypic variables. Finally, we showed that both brain age and the representations underlying it were useful to recognize patients with a range of neuropsychiatric disorders. In paper II we trained further brain age models to investigate its association with genetic variability. Here, we detected seven novel genetic variants associated with differences in apparent brain age, expressed across a variety of tissues. We also performed analyses to investigate the causal relationship between deviations in brain age and neuropsychiatric disorders, without revealing conclusive evidence. In paper III we trained a classifier to differentiate patients with dementia from healthy controls and employed a technique from explainable artificial intelligence to explain its predictions. This allowed us to probe what representations the model had learned, and investigate which regions of the brain it used to make decisions for individual scans. We observed that the model generally focused on regions that are known to contain aberrations in dementia patients and that localizing pathology for individual patients supported prognostic predictions.

Taken together, the current thesis presents evidence that deep learning models can learn to detect complex patterns in neuroimaging data that are associated with aberrations occurring in various neuropsychiatric disorders. Furthermore, we have demonstrated how we can employ a variety of methods to understand what these representations mean, even though the models themselves are incomprehensible. Finally, we have shown that leveraging these representations can elucidate the heterogeneity innate to these disorders and support precise clinical decision-making for individual patients.

2 List of papers

1. **Leonardsen, E. H.**, Peng, H., Kaufmann, T., Agartz, I., Andreassen, O. A., Celius, E. G., Espeseth, T., Harbo, H. F., Høgestøl, E. A., de Lange, A. M., Marquand, A. F., Vidal-Piñeiro, D., Roe, J. M., Selbæk, G., Sørensen, Ø., Smith, S. M., Westlye, L. T., Wolfers, T., Wang, Y. (2022). Deep neural networks learn general and clinically relevant representations of the ageing brain. NeuroImage, 256, 119210. https://doi.org/10.1016/j.neuroimage.2022.119210

2. **Leonardsen, E. H.**, Vidal-Piñeiro, D., Roe, J. M., Frei, O., Shadrin A. A., Iakunchykova O., de Lange, A. M., Kaufmann, T., Taschler, B., Smith, S. M., Andreassen O. A., Wolfers T., Westlye, L. T., Wang, Y. (2023). Genetic architecture of brain age and its causal relations with brain and mental disorders. Molecular Psychiatry, 28, 3111-2310. https://doi.org/10.1038/s41380-023-02087-y

3. **Leonardsen, E. H.**, Persson, K., Grødem, E., Dinsdale, N., Schellhorn, T., Roe, J. M., Vidal-Piñeiro, D., Sørensen, Ø., Kaufmann, T., Westman, E., Marquand, A., Selbæk, G., Andreassen, O. A., Wolfers, T., Westlye, L. T., Wang, Y. for the Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (2023). Characterizing personalized neuropathology in dementia and mild cognitive impairment with explainable artificial intelligence. medRxiv (preprint). https://doi.org/10.1101/2023.06.22.23291592

3 Abbreviations

AI	Artificial intelligence
AD	Alzheimer's disease
AUC	Area under the receiver operating characteristic curve
BAG	Brain age gap
BIP	Bipolar disorder
CNN	Convolutional neural network
DEM	Dementia
DNN	Deep neural network
FDR	False discovery rate
GNN	Graph neural network
GWAS	Genome-wide association study
ICA	Independent component analysis
ILSVRC	Imagenet Large Scale Visual Recognition Challenge
LRP	Layerwise relevance propagation
MAE	Mean absolute error
MCI	Mild cognitive impairment
MDD	Major depressive disorder
ML	Machine learning
MR	Mendelian randomization
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
PAC2019	Predictive analysis challenge 2019
PD	Parkinson's disease
PCA	Principal component analysis
ROI	Region of interest
SCZ	Schizophrenia
SFCN	Simple fully convolutional network
SNP	Single nucleotide polymorphism
ViT	Vision transformer
VGG	Visual Geometry Group
XAI	Explainable artificial intelligence

4 Introduction

In recent years, several neurological and mental, collectively denoted neuropsychiatric (World Health Organization, 2001), disorders have emerged as leading causes of disability and mortality (World Health Organization, 2021, 2022). While these conditions are multifaceted, they share common characteristics such as heterogeneous and overlapping symptomatologies and poorly understood etiologies, giving them an enigmatic appearance. The complexities of their biological underpinnings have fostered interdisciplinary efforts to collect and analyze unprecedented amounts of biological data in search of clues of their true nature. However, identifying patterns in these high-dimensional data require modelling techniques with expressive capabilities beyond what has traditionally been used. The last ten years have seen an unparalleled rise in the popularity of artificial intelligence, where advanced deep learning techniques learn to recognize and exploit complex relationships in data (LeCun et al., 2015). Deep neural networks have displayed their potential by repeatedly outperforming human adversaries in a range of predictive tasks. The maturation of deep learning presents a pivotal opportunity to increase our understanding of the biological aberrations underlying neuropsychiatric disorders, and ultimately alleviate the burden they impose. However, this requires the efforts of experts from various scientific domains to collectively develop and apply the novel methodologies that are needed.

4.1 The neuroscientific perspective on neuropsychiatric disorders

The brain is our most intricate organ and among the most complex structures in the known universe. Through a sophisticated interplay between electrical, biochemical, and physical signalling systems it is the source of our actions and thoughts, hopes and dreams. Ever since Hippocrates, the ancient Greek father of medicine, identified it as the seat of consciousness and the epicentre of perception and action, great minds have been fascinated by the quest to understand how a biological organ can give rise to mental states, a rich inner life, and, ultimately, behaviour. Historically, this investigation has taken on many forms across various scientific disciplines, but since the 1960s many of these have converged under the encompassing umbrella of neuroscience. While the overarching goal of the field still traces its ancient origins, the approaches employed to reach it have changed dramatically, propelled forward by an ever-growing

body of knowledge and technological advances. Nonetheless, despite leaps in understanding, a comprehensive account of the biological substrates of mental phenomena remains elusive.

While an individual mental state is transient, representing an immediate response to internal or external stimuli, they combine in sequences to form longer lasting mental processes. Across these there is an apparent temporal stability to each one of us as individuals, constituting high-level characteristics such as personality traits, cognitive capacities, and ability to regulate our emotions. Aberrations in these composite faculties sometimes lead to an enduring deterioration in the well-being of an individual, or their ability to function in society, constituting characteristic features of a neuropsychiatric condition. This is a broad scope encompassing a wide array of disorders and diseases, but many of them share essential properties. First and foremost, they are clinical syndromes largely recognized and diagnosed with an emphasis on behavioural symptomatology rather than the presence of quantitative biomarkers. This is in part because their aetiologies are largely unknown, indicating that we don't know what causes them to occur in some individuals but not others. Furthermore, the defining clinical characteristics are generally considered dimensional phenomena, where the individuals fulfilling the criteria for a clinical diagnosis experience symptoms on the extreme end of a continuum that also encompasses variability in the general population. Taken together, these defining features portray a complicated clinical landscape.

The scope of disorders that fall within these bounds is extensive, and this thesis will only concern itself with a subset of diagnoses representing some of the diversity it encompasses. Schizophrenia (SCZ) is a mental disorder characterized by a loss of touch with reality manifested through psychotic episodes and hallucinations as well as cognitive and social dysfunction. Bipolar disorder (BIP) entails problems with emotion regulation, with those afflicted alternating between periods of extreme emotional highs and lows. These patients also frequently experience psychotic episodes, most commonly in relation to manic states, the extreme highs. Patients with depression experience low moods and reduced motivation, often combined with impaired cognition. When these symptoms cause a persistent reduction in life quality, this can result in a major depressive disorder (MDD) diagnosis. Dementia (DEM) is an aging-related cognitive condition that is used to describe patients with a cognitive decline severe enough to significantly interfere with their daily lives. In milder cases, where symptoms are apparent but not incapacitating, patients are diagnosed

with mild cognitive impairment (MCI). There are multiple biological aetiologies associated with cognitive decline. The most common is Alzheimer's Disease (AD), recognized by the presence of specific proteins in the brain (Knopman et al., 2021), occurring in approximately 70% of dementia cases. However, it is possible to have dementia without AD pathology, and individuals with AD pathology may display normal cognitive function (Aisen et al., 2017). Multiple sclerosis (MS) is an autoimmune disease where the immune system starts attacking brain tissue. It specifically targets white matter, myelin sheets that wrap nerve fibres to enhance communication in the brain. Over time, this reduces the brain's ability to distribute and process information and decree commands, causing a variety of symptoms related to perception, cognition, and motor control. Symptoms related to the latter are also prevalent in Parkinson's disease (PD), where neuronal atrophy of primarily dopaminergic neurons deteriorates the ability of those afflicted to control their movements.

The above portrayal of the seven clinical conditions included in the current thesis provides a rough nosology describing the subset of neuropsychiatric conditions addressed here. Based on such simplified depictions it is easy to get the impression that these are well-defined, disparate entities. However, this impression does not hold up to scrutiny. SCZ and BIP have a large psychopathological and phenomenological overlap (Pearlson, 2015), highlighting the intricacies of delineating the diagnoses by distinct symptomatologies. This is exacerbated by the existence of comorbidities, when multiple conditions co-occur in the same individual. For instance, patients with SCZ have substantially increased prevalence of MDD (Upthegrove et al., 2017). Furthermore, a variety of cognitive impairments have been observed in SCZ (Kahn et al., 2015; Keefe & Harvey, 2012), BIP (L. J. Robinson & Nicol Ferrier, 2006), and MDD (Pan et al., 2019). Patients with DEM and MCI often display a change in personality, regularly entailing depressive symptoms (Enache et al., 2011). Cognitive impairments are common among patients with MS (Brassington & Marsh, 1998; Risacher & Saykin, 2013) and PD (Emre, 2003; Poewe et al., 2017), and there is a substantial increase in the prevalence of MDD in these two patient groups compared to the general population (Aarsland et al., 2012; Siegert & Abernethy, 2005). Findings like these implicate these conditions as symptomatically overlapping (Woo et al., 2017), high-dimensional (Kupfer & Regier, 2010), and continuous (Helzer et al., 2006; Markon et al., 2011), such that identifying the correct diagnosis can prove challenging (Freedman et al., 2013).

4.1.1 Brain imaging in neuropsychiatric disorders

While the neuropsychiatric disorders outlined above are diagnosed, and often even defined, through the presence of symptoms (Arbabshirani et al., 2017), there is an undeniable concomitance between variability in the brain and along symptomatic spectrums (Bennett & Hacker, 2022). Thus, the neuroscientific scope extends, by encompassing the pursuit to identify and decipher the neural substrates that underlie them.

Since its invention in the 1970s, magnetic resonance imaging (MRI) has provided an unparalleled opportunity to non-invasively study the brain in living organisms. MRI is a versatile medical imaging technique quantifying properties of biological tissue via the application of strong magnetic fields. In both neuroimaging research and clinical neurological examinations, T1-weighted MRI has played a central role in assessing brain structure over the last decades (Symms et al., 2004). Here, a radio frequency pulse is used alongside the magnetic field to rapidly alter the alignment of protons in the atomic nuclei in the brain. When these realign with the magnetic field generated by the magnet, signals are emitted that can be detected by sensors in the scanner. Based on the duration of the realignment, the so-called relaxation time, properties of the tissue surrounding the proton can be derived. In the brain, this tissue is usually classified as either grey matter, mainly comprised of neuronal cell bodies, or white matter, primarily containing nerve fibres wrapped in myelin.

In this thesis, all papers rely on T1-weighted structural MRIs, which from here on will be referred to as structural MRIs. The atomic units in a structural MRI are voxels, the volumetric equivalent of pixels, containing information about a cubical region of the brain with varying dimensionality, commonly $\sim 1mm^3$ (Liang & Lauterbur, 2000). In each voxel, the information harvested by the scanner is encoded as an intensity value, a single numerical value denoting characteristics of the tissue in this circumscribed region of space. This is further interpreted as describing the tissue type located within the voxel, be it white or grey matter, part of the encapsulating skull, or cavities filled with cerebrospinal fluid. When combined, the delineation of tissues between voxels in a structural MRI allows for outlining regions that conjoin to make up the anatomical structure of the brain, hereafter referred to as neuroanatomy.

The information captured in structural MRIs provides a basis for investigating the brains of patients with neuropsychiatric disorders using statistical methods. The high-dimensional information encoded in the images can be condensed

by segmenting anatomical regions of interest (ROIs), a process that was originally performed manually but has later been automatized (Fischl et al., 2002). The volumetric properties of these regions, such as size or shape, can then be compared across patients and healthy controls to detect group-wise differences using standardized statistical methodology. Relatedly, voxel-based morphometry investigates regional brain morphometric properties via differences in individual voxels, as opposed to delineated regions (Ashburner & Friston, 2000). Applying either of these methods to detect differences between individuals assumes spatial correspondence at the relevant measurement level, i.e. that the region or voxel in one individual corresponds to the region or voxel in another individual. This is often achieved through image registration, the spatial consolidation of images against each other or a common template (Jenkinson & Smith, 2001). Subsequently, differences between patients and controls for each measure can be quantified via statistical inference. As such, these approaches are essentially two-step processes, where a preprocessing step produces measures that are subsequently analysed with statistical techniques.

Applications of variants of the analytical approaches outlined above to structural MRI data have yielded a wealth of insights into neuroanatomical aberrations apparent in brain imaging data from neuropsychiatric patients. Case-control comparisons in SCZ have revealed subtle, but widespread, group-level reductions in both white and grey matter volumes in the cortex (van Erp et al., 2018), and subcortically, including smaller hippocampi, amygdala, and thalamus (Kahn et al., 2015). In BIP, reduced cortical thickness has been observed at the group level, most prominently in frontal, temporal, and parietal regions (Hibar et al., 2018). Additionally, reductions in the volumes of subcortical structures have been reported, including the hippocampi and thalamus (Hibar et al., 2016). For patients with MDD, the most consistent finding is that of reduced volume of the hippocampi (Otte et al., 2016; Schmaal et al., 2020), in addition to reduced volumes and surface areas across the cortex (Schmaal et al., 2017). In AD, volumetric reductions in the medial temporal lobe have been recorded years before the onset of clinical symptoms (Knopman et al., 2021). As the disease progresses this neurodegeneration spreads extensively across the brain, a feature shared with the other aetiologies underlying DEM and MCI (Risacher & Saykin, 2013). MS and PD also have a substantial neurodegenerative component as revealed by structural MRIs. In the former, beyond the presence of idiosyncratic lesions that occur where the immune system causes demyelination (Filippi et al., 2018), MRI indices suggesting widespread neurodegeneration have

been reported (Hauser & Oksenberg, 2006). For the latter, the characteristic neuropathological feature is degeneration of the substantia nigra (Poewe et al., 2017), which has been shown to be accompanied by reduced cortical thickness spanning large areas of the cortex (Laansma et al., 2021). Overall, there is clear evidence for the presence of volumetric differences in the brains of patients suffering from neuropsychiatric disorders, most prominently reductions in both cortical and subcortical brain volumes.

4.1.2 Shortcomings of the case-control paradigm in the face of heterogeneity and comorbidity

The literature reviewed above demonstrates that the brain imaging features associated with neuropsychiatric disorders collectively form a complex and heterogeneous pattern. This is further complicated by the discovery that the number of associations reported in the literature as a whole has been shown to dramatically outnumber what is expected statistically (Button et al., 2013; Ioannidis, 2011), indicating that they should be interpreted with care. Relatedly, reported findings of aberrations in patient groups based on neuroimaging data have been notoriously hard to replicate (Boekel et al., 2015; Vogt, 2023). Furthermore, the group-level findings listed above are subject to substantial intra-group heterogeneity, a phenomenon that has been quantified empirically (Wolfers et al., 2018; Wolfers et al., 2021; Young et al., 2018). Many of the aberrations are also overlapping between disorders, underscoring the intricacies of delineating them based on observable differences in neuroimaging data (Venkatraghavan et al., 2023). This has been empirically validated in a meta-analysis across mental disorders (including SCZ, BIP, and MDD) revealing that the available evidence for overlapping volumetric anomalies outweighs that of disease-specific aberrations (Goodkind et al., 2015). In MS, cortical and subcortical grey matter volumes have been shown as better predictors of long-term disability than volumes of the characteristic lesions (Moridi et al., 2022). Post-mortem examinations have found that one third of PD patients with DEM also met neuropathological criteria for an AD diagnosis (Galvin et al., 2006). To conclude, the neuroanatomical viewpoint corroborates the clinical, a continuous landscape where groups are heterogeneous, and their demarcations are obscure. This perspective is also supported by genetic studies, with the discovery of genetic variants that underlie a variety of neuropsychiatric disorders (Brainstorm Consortium et al., 2018).

Even in the instances where the literature is not conflicting, and reported group differences appear stable, it is worth contemplating their implications. Group differences are typically inferred by comparing a group of healthy controls with a group of patients, both including individuals that vary in many respects not encompassed by their diagnostic status (Kopal & Bzdok, 2023). The comparison itself is most commonly operationalized by quantifying the difference between the group means. Taken together, these two properties can yield results that are statistically significant, but where the within-group variability overshadows the between-group difference. This yields discoveries that are plausibly biologically valid but with marginal predictive value for the individual (Bzdok et al., 2020; Davatzikos, 2019). Further aggravating this issue is the impediment that most neuroimaging studies have been carried out in relatively small groups (Marek et al., 2022; S. M. Smith & Nichols, 2018), often combined with suboptimal validation strategies, most prominently in-sample evaluations (Yarkoni & Westfall, 2017). All in all, this has multiple undesirable effects on the knowledge produced by the field. First, inadequate predictive value is limiting in and of itself, as it reduces the translational value of results (Woo et al., 2017). Second, the fact that the knowledge is not predictive implies that the differences that are found fail to encompass the phenomena being investigated. Despite these limitations, descriptive approaches comparing MRI-derived measures between groups using traditional statistical inference remain the predominant methodology in the field (Loth et al., 2021).

4.1.3 From description to prediction: New venues for modelling the relationship between brain and behaviour

To address the challenges presented by group-level, descriptive approaches, new venues have been proposed over the last decade, aiming to increase the fidelity of the models depicting the relationship between neuroimaging features and behaviour. One is the reconceptualization of central ontological ideas. This is epitomized through the reconsideration of the suitability of the case-control paradigm with its implication of two disjunct, homogeneous, groups, an idealization that does not appear to represent clinical nor biological reality in neuropsychiatric disorders. Instead, new conceptual approaches conceive patients as a heterogeneous group that can deviate from the normalcy represented by the controls in various ways (Marquand et al., 2019). One instance of such is normative modelling (Marquand et al., 2016), where this overall goal is opera-

tionalized through methodology that resembles pediatric growth charting. Here, data is used to fit distributions for any given neuroimaging measure conditioned on other variables, typically age, to outline a trajectory of values that are considered normal (Bethlehem et al., 2022; Rutherford et al., 2022). Normative modelling has been used to model neuroanatomical aberrations in SCZ, BIP, and MDD (Segal et al., 2023; Wolfers et al., 2018), and characterize diversity in DEM (Verdi et al., 2021), proving its efficacy as a conceptual framework for embracing heterogeneity.

Another emerging perspective in psychology and clinical neuroscience the last decade has been a shift from description to prediction (Bzdok et al., 2021; Rosenberg et al., 2018; Yarkoni & Westfall, 2017), made possible by machine learning (ML) methods. A fundamental difference between the traditional, descriptive framework and ML is what kind of knowledge the two methodologies aim to produce (Breiman, 2001). In the simplistic case-control setting, the former aims to identify and describe differences between the groups as outlined above, typically resulting in a set of measures where the groups significantly differ. Conversely, predictive ML aims to identify which participants belong to which group. The first step towards this goal is to procure a singular prediction per individual. In its simplest form this is done using multivariate statistics, where group-wise differences for each measure are identified, equivalent to the descriptive approach. However, this is taken a step further, by comparing each separate individual with the group-wise prototypes defined through these differences and collapsing the results into a single prediction. Thus, instead of procuring spatial maps of the brain highlighting locations where the groups differ, predictive studies are evaluated through predictive performance measures indicating how well they classify individuals, such as accuracy or mean prediction error (Varoquaux & Colliot, 2023). Furthermore, this general shift towards prediction is often coupled with an emphasis on out-of-sample generalization as opposed to in-sample characterization as the main validation of the utility of models (Bzdok, 2017), which is important to assess the translational potential residing in clinical neuroimaging (Woo et al., 2017). The focus on generalization and rigorous validation is also a natural remedy to counteract the ongoing replication crisis (Botvinik-Nezer & Wager, 2022). More generally, a shift towards prediction has the potential to profoundly change the mindset of researchers, maturing clinical neuroscience as a quantitative, empirical field of research (Yarkoni & Westfall, 2017).

The emphasis on predictive performance is naturally accompanied by the

introduction of more expressive and accurate methods to model the relationship between brain and behaviour (Bzdok & Yeo, 2017). Why the former enables the latter is worth elaborating. One of the main benefits of the conventional descriptive approach is its innate interpretability. Group-wise differences are found in circumscribed spatial locations, either predefined ROIs or voxels, that facilitate biological interpretations. This remains true also in the simplistic multivariate predictive context outlined above. However, this approach also has innate limitations. First, it necessitates a practical framework for ensuring that the comparisons between individuals are based on comparable units, either by forcing the images into the same stereotactic space or extracting ROIs. This places a heavy burden on the veracity of the preprocessing methods underlying this step (Bookstein, 2001). Second, it limits information being shared beyond these units, obstructing the possibility of detecting patterns where information in different spatial locations is combined. This imposes a strict restriction on the types of patterns that can be discovered and utilized to differentiate the groups in the modelling process, limiting the scope of phenomena the models are able to capture. Specifically, these types of modelling pipelines will be able to separate cases from controls to the degree made possible by a linear separation in either voxel or ROI space, a conceptual level that has been argued not to be sufficient to discern the aberrations underlying neuropsychiatric conditions (Davatzikos, 2004). This has fuelled a quest for more intricate modelling machinery able to apply complex pattern-matching strategies to increase predictive performance.

The role filled by preprocessing software extracting volumetric properties of specific ROIs is to exploit information from the raw voxel space of an MRI volume to procure higher-level idealizations that can be used for modelling. This process of selectively aggregating information into a generalized, conceptual form is known as forming an abstraction. Opposed to the high-dimensional continuous data in a raw MRI volume, the level of abstraction represented by volumetric properties plays right into our plausibly evolutionarily beneficial propensity for seeking to describe the world in terms of delineated objects and their characteristics (DiCarlo et al., 2012). However, there is no principled reason to believe this is the correct level of abstraction for detecting canonical neuroanatomical patterns to characterize complex neuropsychiatric disorders, if they exist at all. Contrarily, there is evidence that the processing streams in the brain responsible for higher cognition are widely distributed among non-adjacent locations (Westlin et al., 2023), implying that any structural brain aberration that may cause their dysfunction could be of a similar nature (For-

nito et al., 2015). As modelling phenomena at an appropriate level of abstraction is of the utmost importance (Figure 1), this has spurred a quest for finding modelling techniques that appropriately match the level of abstraction where deviations plausibly emerge. In its simplest form, this has been done by retaining the abstraction level achieved by preprocessing tools before modelling but introducing more complex statistical machinery to perform the modelling itself (Mateos-Pérez et al., 2018). In neuroimaging, the early classes of models commonly adopted were support vector machines (Arbabshirani et al., 2017; Wolfers et al., 2015) and random forests (Davatzikos, 2019), both more expressive than regular linear models. A complementary approach has been to extract useful data-driven abstractions prior to modelling, most commonly using either principal or independent component analysis (PCA and ICA, respectively), or the closely related partial least squares (Mwangi et al., 2014). However, while successfully applied in several contexts, neither of these caused a paradigmatic shift in the collective ability to predict clinical or behavioural phenotypes based on neuroimaging data.

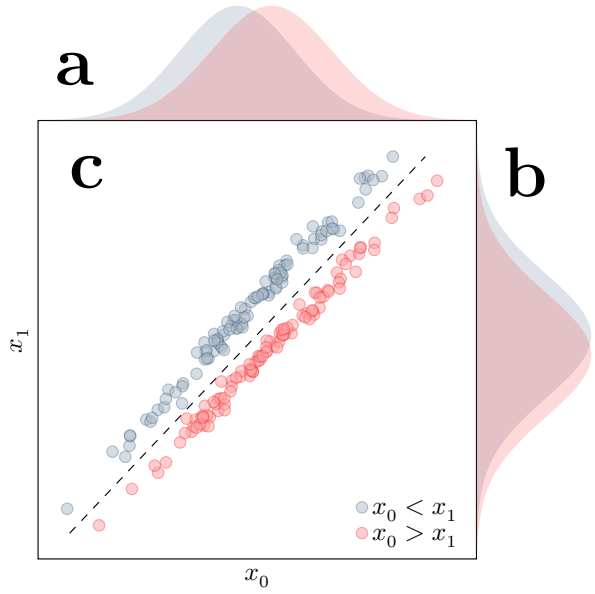


Figure 1: Two sets of data points seen from different perspectives. (a) and (b) When the distribution of values in the two groups are assessed with respect to the two dimensions independently, the group means are different, but with substantial overlaps. (c) When the points are seen in two-dimensional space, the groups are linearly separable.

The search for appropriate abstractions in neuroimaging has also been pursued in a different direction, inspired by knowledge about the structure of the brain. On an abstract level, the brain can be conceptualized as a graph of interconnected neurons (Mitchell, 2009), where information processing is distributed and parallelized. One way of describing the overall characteristics of a graph is through its topological structure, broadly denoting patterns of interconnections between its nodes. The topology of the brain has been described as hierarchical (Sporns, 2013), containing neurons that cluster to form functionally or anatomically disparate regions (Glasser et al., 2016), interconnecting form circuits (Tau & Peterson, 2010), which again conjoin to form networks (Thomas Yeo et al., 2011). Theoretically, a complex system with hierarchical properties can be modelled via an analogous hierarchy of abstractions (Floridi, 2013). In the latter, each level in the hierarchy corresponds to a level of abstraction, such that the bottom level describes the system at full resolution, whereas the higher levels successively condense salient information into compound representations. In this idealized, conceptual model, it is possible that information that appears scattered and unorganized at lower levels can be combined into meaningful patterns at higher levels.

Current MRI technology does not provide the spatial resolution necessary to accurately disentangle the interconnections between neurons. However, the combination of structural and other MRI modalities has been used to describe interconnections between brain regions more coarsely (E. C. Robinson et al., 2010). Based on this a high-level graph structure can be imposed, which can then form a basis for modelling neuroimaging data at various levels of abstraction. This has recently been combined with normative modelling to procure promising results for unifying neuroanatomical deviations in heterogeneous patient groups at different levels (Segal et al., 2023). However, this general approach has also faced criticism: the exact levels of abstraction used in a given analysis rely on the methodology used to identify the graph structure, leaving room for potentially arbitrary choices of analytical strategy (Botvinik-Nezer et al., 2020). Subsequently, this has yielded results that are hard to replicate and have limited translational value (Bijsterbosch et al., 2020). Moving forward, it will be beneficial to explore approaches for modelling heterogeneity in the brain based on abstractions that are rooted in data, to complement those relying on human efforts to uncover useful structures.

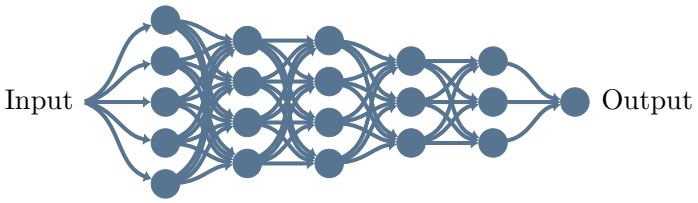


Figure 2: A deep, artificial neural network, where the relationship between inputs and outputs are modelled by layers of artificial neurons.

4.2 Representation learning and image recognition

While the notion of an abstraction is intrinsically vague, formal roots in statistical learning theory offer a conceptual structure and precise nomenclature to describe it. In traditional statistical modelling, the process of combining information from different variables to a meaningful abstract composition happens either explicitly through an interaction term, or *a priori* feature engineering (Zheng & Casari, 2018). Here, handcrafting features, exemplified in neuroimaging by the volumetric features procured by an MRI preprocessing pipeline, is considered a vital part of the modelling process. Yet, it is also recognized as a labor-intensive endeavour with results largely reliant on the human engineer's imaginative bounds. Hence, new venues for automatically learning more effective abstractions have been explored, culminating with the construction of the Deep Belief Network (Hinton et al., 2006). This was an early deep neural network (DNN), a multi-layered variant of the more general class of statistical learning models called artificial neural networks. These model the relationship between input variables and outputs with artificial neurons, simple computational units connected in sequential layers to form a computational graph (Figure 2). The fundamental idea underlying the architecture is that each layer assimilates and refines information from its predecessor, effectively learning to combine data into a hierarchy of representations (Hinton, 2007), giving birth to the notion of representation learning. These learned representations correspond to various non-linear combinations of the input variables, fulfilling the goal of automatically learning abstractions from data. The benefit of this modelling approach is a massive increase in expressiveness, allowing the models to represent far more complex relationships between inputs and outputs. On the other hand, its downside is that the rules describing these relationships become harder

to decode and understand, giving rise to the notion of DNNs as black boxes unintelligible to humans. Nonetheless, the introduction of representation learning operationalized through deep neural networks, combined with novel hardware and the accumulation of vast amounts of data, set the stage for the 21st-century renaissance for artificial intelligence (AI).

The first field to be revolutionized by DNNs was computer vision, where applications of the technology quickly matured to a level yielding practical value, by matching, or even surpassing, human performance at a range of visual tasks. A catalyzing factor underlying these advances was the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), a yearly modelling competition put forth to motivate innovation and provide an objective measure of the prevailing state-of-the-art (Russakovsky et al., 2015). Here, combatants were tasked with automatically classifying samples from the ImageNet dataset, a database containing millions of images from thousands of classes spanning a multitude of domains (Deng et al., 2009). Over the eight years the competition ran the error rate was sliced from 28.2% in 2010 to 2.25% in 2017, more than halving the 5.1% achieved by human antagonists. The largest year-on-year improvement occurred in 2012 (Figure 3) with the introduction of convolutional neural networks (CNNs), a class of deep learning architectures tailored specifically for image data (LeCun et al., 1989). Inspired by the mammalian visual cortex, a part of the brain processing visual information, these models solve predictive tasks by employing a hierarchy of pattern detectors trained to recognize patterns at various scales and levels of abstraction (Goodfellow et al., 2016). These pattern detectors are implemented in artificial neurons, structured in such a way that those early in the model learn to recognize simple visual patterns, which are combined by sub-

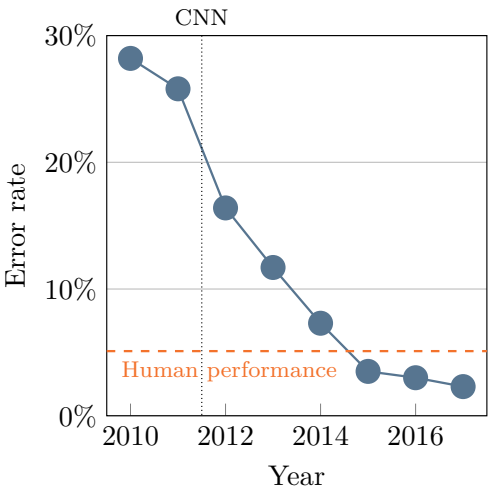


Figure 3: Predictive performance of the winning team in the Imagenet Large Scale Visual Recognition Challenge. The dotted line indicates the pivotal moment when convolutional neural networks were introduced.

sequent ones into more complex concepts. Through this hierarchy, the model effectively learns a conceptual taxonomy representing visual objects and their interrelations, sufficient to solve the task at hand. The vast capabilities of CNNs to learn complex visual representations are most often solely implied through their predictive efficacy. However, researchers have also probed the representational space internal to such models, to explicitly investigate what they have learned (Zeune et al., 2020). Here, the evidence suggests that the representations learned by CNNs can be semantically meaningful (Yosinski et al., 2015; Zeiler & Fergus, 2013; Zhou et al., 2015), corresponding to concepts and objects that are sensible also to humans (López-Rubio, 2021). This gives hope for the utility of CNNs as epistemic tools: efficient representation learners that can manoeuvre complexities in data to solve difficult tasks, potentially in ways that are both informative and comprehensible to the human eye (Durstewitz et al., 2019).

4.2.1 Deep learning in neuroimaging

The neuroimaging community quickly embraced deep learning (Plis et al., 2014) as a set of efficacious predictive tools with two distinct goals. First, DNNs have been adopted as means to optimize isolated steps in a standard neuroimaging processing pipeline, to procure ideal images for subsequent analysis. A prime example of such is Fastsurfer (Henschel et al., 2020), a deep learning-based tool for brain segmentation, built to perform the same tasks as previous tooling with comparable results in a fraction of the time (Bloch & Friedrich, 2021). Secondarily, deep learning models have been used as end-to-end models to predict outcomes based on raw imaging data, such as clinical diagnoses or other phenotypical variables (Noor et al., 2020; Zhang et al., 2020). Here, their capabilities as representation learners are put to full use, as working with the images themselves does not restrict them to features engineered through other processes (Abrol et al., 2021). This also alleviates the human and computational strain to procure these features (Davatzikos, 2019) and allows for the discovery of patterns that could have been overlooked by human counterparts. Consequently, there have been reports of a general increase in predictive efficacy using end-to-end CNNs (Abrol et al., 2021; Quaak et al., 2021). This can be interpreted as early evidence supporting the innate suitability of DNNs to model brain imaging data: their hierarchical nature could prove ideal for detecting nuanced, spatially distributed patterns at multiple scales, potentially detecting

the aberrations underlying neuropsychiatric disorders (Plis et al., 2014).

When applied directly to classify patient groups based on structural MRIs, three-dimensional CNNs have shown promising results (Figure 4). Combining results from 76 studies compiled in two recent reviews revealed a mean accuracy of 93.06% for models differentiating AD patients from healthy controls (Mirzaei & Adeli, 2022; Sharma et al., 2023). Similarly, compiling results from 56 studies classifying MCI patients resulted in a mean accuracy of 86.99% (Fathi et al., 2022; Sharma et al., 2023). For MS, with its characteristic brain lesions, end-to-end modelling is sometimes replaced with a two-step process of first identifying or segmenting lesions, followed by predicting disease status. A review including both approaches identified 12 studies with a mean accuracy of 89.82% (Shoeibi et al., 2021). Further examinations of recent reviews revealed similar results in PD, with a mean accuracy of 86.85% aggregated across 10 studies (Mei et al., 2021; Noor et al., 2020), and SCZ, with 84.43% from 18 studies (de Filippis et al., 2019; Quaak et al., 2021; Verma et al., 2023). However, it is worth noting that the three latest results are based on fewer studies with substantially more variability. There are, to the author's best knowledge, no reviews on classification in either BIP or MDD that include more than one deep learning model trained on structural MRI data published to date. Investigations into singular studies revealed accuracies of approximately 80% for both the former (Martyn et al., 2019; Saghayan et al., 2023) and the latter (Gao et al., 2023; Hong et al., 2022; Korda et al., 2021; Y. Wang et al., 2021). While altogether promising, it is worth contextualizing these results somewhat. Concerningly, most of them are achieved in very small datasets, and the performance of similar classifiers has been observed to decrease as the datasets grow (Arbabshirani et al., 2017; Janssen et al., 2018; Wolfers et al., 2015). Furthermore, they employ a wide variety of validation approaches, and there is a large variability in the results that have been achieved. Both in the list above and elsewhere, there have been studies reporting accuracies of 100%, an outcome that has been deemed improbable (Woo et al., 2017). These factors combined indicate that the results should be interpreted with caution. It should also be mentioned that similar performances have been reported for simpler modelling frameworks. Although direct comparisons across studies are complicated (Arbabshirani et al., 2017), some studies have performed them, finding evidence both for (Quaak et al., 2021) and against (Moazami et al., 2021) the superiority of deep learning for predicting case-control status from neuroimaging data. Nonetheless, these early results represent a hopeful foundation for the successful adoption of deep

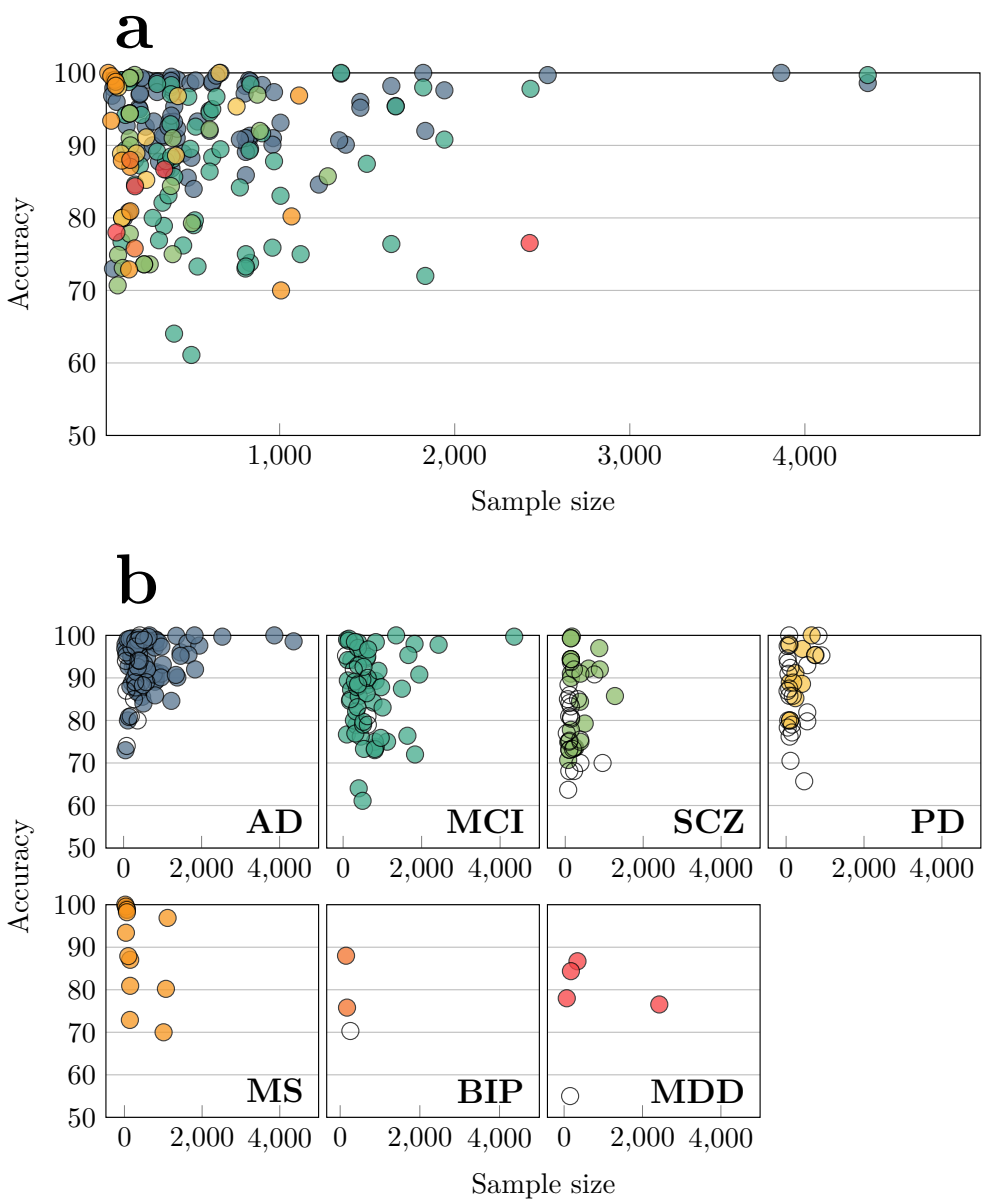


Figure 4: Accuracies achieved by the case-control classifiers compiled from the publications in Section 4.2.1. (a) Results for all deep learning studies across all disorders. (b) Deep learning vs traditional machine learning for each independent disorder.

learning in clinical neuroimaging.

While the predictive performances of CNNs used for case-control differentiation are encouraging, this general approach has also been subject to criticism. For instance, it continues to perpetuate the dichotomous paradigm, rather than exploiting the expressive potential of these models to characterize heterogeneity in patient groups. Furthermore, it builds upon the simplified notion of patients including only those that already have been identified using other means (Davatzikos, 2019), restricting its clinical utility. Their epistemic utility is also limited: although the predictive efficacy of a model might be excellent, its black-box nature makes it hard to derive scientific knowledge from, beyond *quod erod demonstrandum*-type evidence that the imaging data contains information sufficient to differentiate the groups to some degree. To overcome this limitation, researchers have tried to extract more subtle, nuanced knowledge by attempting to understand what the models have learned. One approach to reach this goal is to look at the computation, and particularly the activation of the representations, that happen internally in the model when it sees an image. As these internal representations correspond to visual concepts (López-Rubio, 2021), this allows us to infer to what degree a specific image relates to each representation. When done across multiple MRIs, we can further deduce whether there is structure in this representational space that is comprehensible, indicating that the representations themselves correspond to concepts that appear meaningful. An application of this technique to a CNN trained on structural MRIs revealed a representational landscape that was more discriminative with regards to the phenotypes of interest, than those learned by simpler models (Abrol et al., 2021). This corroborated early evidence of the same phenomena in substantially simpler CNNs for case-control classification (Plis et al., 2014). Applied to functional MRI this technique has revealed representational spaces that covary with abstract clinical and behavioural measures and psychological constructs beyond what was used for training the model (Zabihi et al., 2021). These findings give hope that the representations learned by CNNs can help discern brain heterogeneity related to interesting phenotypes. However, to utilize them towards epistemic and clinical goals, we should pursue understanding what they encode.

4.2.2 Brain age and proxy measures

The simplest way to ensure that a CNN learns an understandable representation is to optimize for that goal directly. This is trivially achieved by using a variable encoding the representation as the predictive target, with disease classifiers being typical examples. However, a major impediment to applying deep learning models to clinical neuroimaging problems is the scarcity of data (Arbabshirani et al., 2017). This is further exacerbated in the case of data with clinical phenotypes (Horien et al., 2021), the typical outputs of interest. To alleviate this problem innovative approaches leveraging ongoing large-scale efforts to collect population data are proposed. One of these is to first learn to model variability inherent in large healthy cohorts concerning common variables, broadly called proxy variables (Dadi et al., 2021), before subsequently applying the models to discern clinical phenotypes. The overall approach of learning and applying a normative reference curve to quantify aberrations is closely related to normative modelling, which has shown promising for disentangling the heterogeneity innate to neuropsychiatric patients (Marquand et al., 2019). Furthermore, it rings of the deep learning mantra "representation first", emphasizing the importance of first learning to identify patterns in data that describe variability, potentially at a high level of abstraction, before employing these to solve a task (Goodfellow et al., 2016). The combination of these two features could prove potent to discern brain heterogeneity related to neuropsychiatric disorders.

The most prominent example of a proxy measure is the approach aptly referred to as brain age, where a statistical model learns to predict the age of a brain based on its visual appearance. The concept of brain age can be described from multiple perspectives. First, it is a normative statistical approach, where a model learns to recognize the average brain across the lifespan based on a training dataset (Cole & Franke, 2017). Subsequently, new brains can be placed in reference to this norm (Figure 5), by comparing the age predicted by the model with the chronological age of the individual (S. M. Smith et al., 2019). This produces a measure of whether the brain as a whole appears younger or older than counterparts in the population data of the same chronological age (Cole et al., 2019). The introduction of aging provides a natural segue to the perspective of biological aging. Here, the apparent age of the brain can be interpreted as a measure of the accumulation of cellular damage or "wear-and-tear" that occurs during a lifespan (Ferrucci et al., 2020; Hayflick, 2007), and whether this has accrued faster or slower than expected. Finally comes the

perspective of brain age as an information theoretical construct, where all the high-dimensional information in a brain scan is compressed into a single number. To summarize, brain age is intuitive due to its statistical properties, even when the underlying computation is unknown. It is related to health, due to its sensitivity towards biological processes. And finally, it is summarizing, due to its computational nature. These three properties combined outline a promising biomarker for brain health.

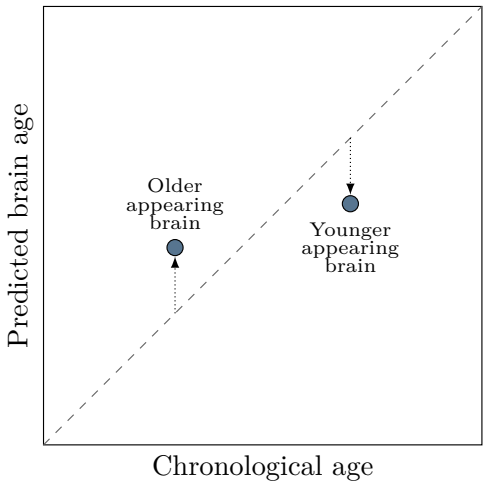


Figure 5: Predictions from a brain age model can indicate whether a brain appears older or younger than a reference norm.

The first model for brain age estimation was fit using relevance vector machines on principal components derived from heavily processed T1-weighted structural MRIs (Franke et al., 2010), and as such represents the traditional predictive framework based on feature engineering and models with restricted expressive capabilities. Nonetheless, the model was able to learn visual patterns coherent with age, predicting brain age with a mean absolute error (MAE) of 5 years. Furthermore, the residuals of the model, encoding the difference between predicted and chronological age (the brain age gap, abbreviated to brainAGE by the original authors, also referred to as the brain age delta), was significantly higher in a group of patients with mild AD compared to healthy controls (Franke & Gaser, 2012), early evidence for its sensitivity towards neuropsychiatric disorders. Since then, this initial finding has been given credence by a wealth of empirical studies using different statistical methodologies to uncover similar patterns across different disorders and datasets (Franke & Gaser, 2019). The observation of elevated brain age has been replicated numerous times in both AD (Beheshti et al., 2019; Löwe et al., 2016; Yin et al., 2023) and general DEM (Kaufmann et al., 2019; Persson et al., 2023; J. Wang et al., 2019). A similar, but smaller, deviation has been reported in MCI (Franke & Gaser, 2012; Kaufmann et al., 2019; Löwe et al., 2016; Yin et al., 2023), and substantially increased brain age has been reported repeatedly in patients with both MS (Cole et al., 2020; Høgestøl et al.,

2019; Kaufmann et al., 2019) and PD (Beheshti et al., 2019; C. R. Eickhoff et al., 2021). A potentially more surprising finding is that of a higher brain age in patients with SCZ (Hajek et al., 2019; Kaufmann et al., 2019; Koutsouleris et al., 2014; Nenadić et al., 2017; Schnack et al., 2016; Shahab et al., 2019) and MDD (Han et al., 2020), although commonly less pronounced and with more variability. For BIP the results are unclear, with studies reporting both increases (Kaufmann et al., 2019) and no significant differences (Hajek et al., 2019; Nenadić et al., 2017; Shahab et al., 2019) in brain age when compared to healthy controls. Finally, as an ultimate testimony of its ability to encode generalized brain health, higher brain age has been associated with higher mortality (Cole et al., 2018). Through these empirical findings, the value of brain age as a biomarker of general, but unspecific, brain health is supported, enabling it both as a potential clinical tool and as a composite, dimensional measure that can further our understanding of the complex neuroanatomy underpinning neuropsychiatric disorders.

Beyond its efficacy for differentiating cases and controls, brain age offers a multifaceted lens through which the intricate interplay between biology, lifestyle, cognition, and psychology can be elucidated. Here, it plays the role of an intermediate phenotype seeking to connect low-level biology with high-level psychological constructs and behaviour. To establish these connections, studies have combined theory-driven pursuits with broad explorations to reveal a wealth of potential interconnections. These have revealed genetic variants correlated with relative differences in brain age (Jonsson et al., 2019; Ning et al., 2020; S. M. Smith et al., 2020), forming a plausible basis for dissimilarities between individuals. Associations with higher-level biology have also been established, both by correlating it with physical and health-related measures (Beck et al., 2021; S. M. Smith et al., 2020), and by relating it to other measures of biological ageing (Elliott et al., 2019), physical health (Cole et al., 2018), and self-reported measures of biological maturity in youth (Holm et al., 2023). In terms of lifestyle factors, relative differences in brain age have been associated with socioeconomic status (Busby et al., 2023; Cohen et al., 2023), and elevations have been observed in those with high alcohol intake (Cole, 2020; Franke et al., 2013; Ning et al., 2020) and smokers (Bittner et al., 2021; Cole, 2020; Linli et al., 2022). Lastly, correlations have been discovered with psychosocial constructs such as well-being (Korbmacher, Gurholt, et al., 2023; Sone et al., 2022) and a subjective experience of own age (Kwak et al., 2018). Taken together, these findings elucidate the integrative role of brain age, associated with both biological measures, lifestyle

factors, and psychological constructs.

Despite these promising empirical discoveries, brain age has also been subject to criticism. Ontologically, it is hard to determine exactly what the measure means, beyond its conceptual association with age as operationalized through statistical modelling (Butler et al., 2021). It can be complicated to determine exactly what contributes to the prediction of the model (Cole & Franke, 2017; Tanveer et al., 2023). This leads to uncertainty about whether it mostly encodes differences that are stable across the lifespan, or if it meaningfully captures ongoing change (Vidal-Pineiro et al., 2021), for instance related to current pathological processes. Clinically, the utility is hampered by within-subject variability, which becomes apparent through predictions that vary with a magnitude of years based on scans from the same individual taken days apart (Korbmacher, Wang, et al., 2023), exacerbated by scanner differences (Baecker et al., 2021). Thus, it is hard to imagine brain age as an imaging biomarker for clinical use in its current state, and further work is required to enhance its technical robustness, clinical validity, and reliability.

4.2.3 Understanding deep neural networks with explainable artificial intelligence

Using deep learning for brain age modelling can be conceptualized as a variant of representation learning where an intermediate phenotype, the type of construct that would typically emerge inside the model, is instead explicitly modelled as the target of the modelling procedure. If this phenotype is truly intermediate, the learned representation, now corresponding with the output of the model, would necessarily need to be operationalized through successive computational steps to contribute towards the actual end goal, for instance predicting case-control status. The main benefit of this two-step approach is that the intermediate phenotype will necessarily be more understandable than if it occurred deep within the model. A potential shortcoming is that although the intermediate representation is understandable, it is not necessarily useful. To avoid this, it would be better to construct a model that solves the task of interest, with full expressive freedom to construct the representations that are needed, and then retrospectively figure out what those representations mean. Predictive modelling using deep learning attempts to solve the first half of this problem, but resolving the second has proven difficult.

Once trained, all DNNs can be written out as a concrete mathematical for-

mula, where the representations it has learned, and their interrelations, will be fully defined. In a CNN trained on structural MRIs, this math will roughly correspond to a hierarchy of pattern detectors (Balestriero & baraniuk, 2018), each detecting a three-dimensional visual concept related to brain anatomy (Khorrami et al., 2015). This mathematical formula is fully accessible; however, it will necessarily encode a highly complex, non-linear mapping between MRI images and predictions, utterly incomprehensible to humans (Samek & Müller, 2019). To grasp the true meaning of the representations they have learned we need to associate them with familiar concepts and terminology, to foster a holistic understanding of how they arose and what they encode (Cappelen & Dever, 2021). However, this is no simple task. DNNs fall within the scope of connectionist models (Goodfellow et al., 2016), where a multitude of computational units solve problems collectively via propagating continuous values among themselves (LeCun et al., 2015). Somewhat surprisingly, as these are inspired by the inner workings of our brains, this gives rise to a mode of reasoning that does not appear intuitive to us. Where we normally think in terms of symbols, discrete representations that correspond to specific and delineated concepts and objects, and their interrelation, DNNs rely on sub-symbolic representations. This entails internal representations that encode information by distributing it across vast vectors of continuous numbers (Calegari et al., 2020), completely circumventing the notion of well-defined symbols. Whether these two operational modes can be reconciled remains an open philosophical question (Smolensky, 1987).

A more practical approach towards bridging the apparent gap between the decision processes of DNNs and humans is taken through explainable AI (XAI), where concrete methodology is offered to demystify the inner workings of the former. While the nomenclature varies (Barredo Arrieta et al., 2020), a common formulation of its defining purpose is to create methods that explain decisions made by a DNN in a human-understandable manner. A crucial nuance in this definition is that the quality of an explanation is not governed by universal law, but rather depends on its recipient's capacity for understanding, for instance relying on their preexisting knowledge about the domain at hand. XAI is becoming a valuable tool to safeguard that a model has learned to solve a problem in a fashion that is recognizable to a human domain expert, fostering trust in the model's reliability and generalizability. In safety-critical domains, including medicine, lack of trust has been pointed out as a main culprit underlying the skepticism towards the practical adoption of AI (Kundu, 2021), emphasizing the translational potential of XAI. For real-life clinical scenarios, the most com-

mon use case proposed for XAI is in tools providing clinical decision support. If successful, this would allow a highly efficacious predictive model to operate in tandem with a clinician with deep domain expertise, supplementing human intuition with immense number-crunching capabilities and an eye for detail, to provide the best possible care to those in need.

There exists a variety of different XAI techniques, applicable in various contexts depending on the model that is to be explained, who the explanation is for, and what it wants to achieve (Samek & Müller, 2019). Most common for DNNs are post hoc techniques that provide an explanation for each individual prediction of the model. In the case of CNNs, these explanations are often visual, comprised of a heatmap indicating which regions of the image contributed to the prediction of the model (Simonyan et al., 2014). Multiple heatmapping techniques exist, mostly differing in their notion of what constitutes a contribution. Early methods, starting with saliency mapping, relied on the backpropagation of gradients to highlight contributing regions. An innate issue with these approaches is their literal interpretations: a highlighted region contains "what it should contain even more of for the prediction to be even higher". Although this is often more intuitive in practice than in theory, this somewhat convoluted definition has led to the development of alternative methods. An example is Layerwise Relevance Propagation (LRP), where regions are instead highlighted based on how much they contribute to the prediction (Bach et al., 2015). The term relevance comes from the unit that is used to denote contribution, and the produced heatmaps are aptly called relevance maps. A relevance map is generated by propagating relevance backward from the output space of the model to the input space, effectively smearing the prediction out across the pixels (or voxels in the three-dimensional case) of the input image (Figure 6). Thus, the relevance that ends up in a specific location denotes how much the pixel (or voxel) in that exact position contributed to the prediction. Since its invention, multiple variants of LRP have surfaced, created to emphasize different aspects of the decision process (Montavon et al., 2019). These can also be combined, to create even more elaborate explanation schemes, proving more useful than their simpler predecessors (Kohlbrenner et al., 2020).

Although the visual explanations obtained via heatmapping techniques appear meaningful, this does not necessarily entail they are veracious renditions of anything related to the decision process that occurs within a model. Investigations have revealed that these methods appear more sensitive towards visually salient features of the image, such as sharp edges, as opposed to information

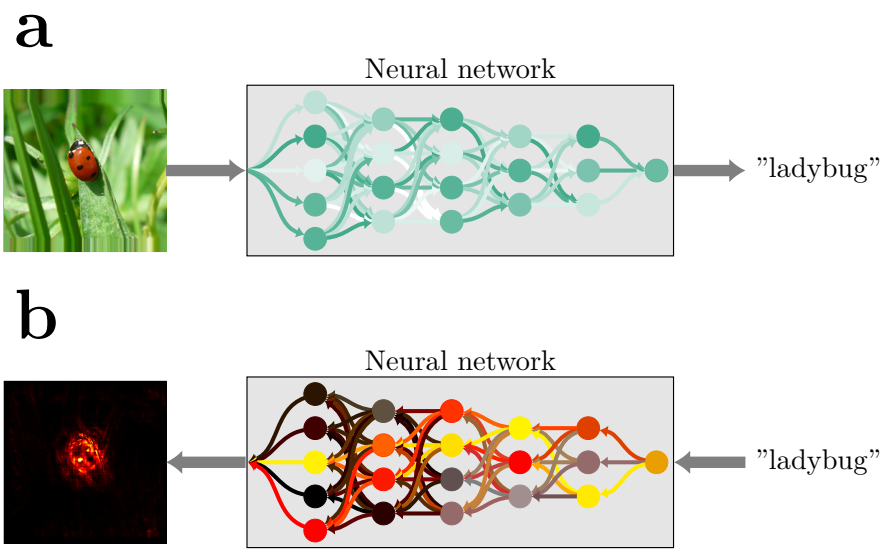


Figure 6: The application of Layerwise Relevance Propagation (LRP) to a very simplified neural network. (a) Through a forward-pass the model predicts that the image contains a ladybug. (b) Through a backward-pass LRP computes a heatmap indicating which regions of the image contributed to the prediction.

that is useful to determine its content, plausibly used by the model (Adebayo et al., 2020). Furthermore, it has been shown that explanations produced to explain different predictions from the same model are often qualitatively similar (Sixt et al., 2020). This is exemplified by the scenario when a classifier is given an image containing both a cat and a dog and tasked to predict the probability of the image containing each animal independently. When heatmapping techniques are used to explain the two predictions, the resulting explanations are indistinguishable. Finally, it has been reported that large, qualitative, changes to an image do not impact the explanation that is generated (Kindermans et al., 2019). Altogether, this implies that these techniques should be used with care. Furthermore, as the degree to which these failure modes occur depends on the specific model, task, and heatmapping technique used, they highlight the importance of validating that the explanations procured are truthful and informative. However, as there rarely exists ground truth labels of why an image contains what it contains, and how the model should recognize this, this can be a difficult task.

XAI has not yet found a foothold in clinical environments relying on neuroimaging data. However, the empirical groundwork required to support this endeavour is being rapidly built by the research community. In DEM, an exponential increase in the number of papers applying XAI has been observed over the last few years (Martin et al., 2023). Most of these rely on imaging data and CNNs and use variants of heatmapping techniques, often LRP, to provide post-hoc explanations of individual predictions. The explanations generated in such studies have been observed to corroborate existing knowledge of neuropathology (D. Wang et al., 2023), giving credence to the notion that the classifiers are accruing relevant knowledge. Through discoveries like these, studies on DEM and AD (Böhle et al., 2019; Dyrba et al., 2021) have spearheaded the effort of showing the potential for XAI as an asset for analysing clinical neuroimaging data (Farahani et al., 2022; Rahman et al., 2023), with other neuropsychiatric disorders following suit. In MS, LRP has been used to show that a CNN trained to recognize patients uses both lesions, the most prototypical biomarker for the disorder, and other sources of information to perform its task (Eitel et al., 2019). For MDD, XAI based on functional MRI data has been used to identify imaging features that are associated with treatment response (Squires et al., 2023). In PD (Camacho et al., 2023; Magesh et al., 2020), SCZ (Lin et al., 2022; Rahman et al., 2022), and BD (Saglam et al., 2023), various XAI techniques applied to different imaging modalities have implicated brain regions affected by the dis-

ease. Recent years have seen multiple arguments for the immense potential of XAI to support clinical practice, also in neuropsychiatric disorders, but the field is still in its infancy, and the literature is sparse. To reach its potential, empirical studies are imperative to identify how, when, and where these techniques add value.

XAI was born from a practical desire to understand whether a DNN was solving its task in a reasonable manner, articulated from the point of view of the developer or the user of an AI system, wanting to ensure its proper functioning. However, it also has the potential to provide a powerful epistemic tool, that allows us to understand sophisticated inferential patterns at the heart of complicated phenomena that would otherwise be beyond our cognitive grasp. This is especially compelling in the case of neuropsychiatric disorders because of their enigmatic appearance and apparent biological complexity. However, whether the process of fitting a complex, expressive, DNN to data, and then applying post-hoc techniques for understanding how it works, is a viable strategy for attaining scientific knowledge is an open question. It has been argued that explainability is mostly useful for a specific stakeholder to understand a specific model, not to to produce facts (Páez, 2019). Determining what these approaches can be used for relies largely on philosophical standpoints, epistemic beliefs, and assertions about what these models are and what they do. Despite the abstract nature of these topics, their resolution will be supported by empirical data, adding incentives to continue exploring the utility of XAI in clinical neuroscience. Whether AI, and particularly the explainable kind, will help unravel the mysteries of the mind, is a question for the future.

4.3 Validating imaging discoveries with genetic analyses

As outlined so far, intermediate phenotypes describing variation in the brain, including abstract ones learned by advanced deep learning models, have the potential to help us disentangle neuroanatomical heterogeneity in the context of neuropsychiatric disorders. However, these are composite measures that will have causes of their own, an understanding of which would even further elucidate the association between brain and behaviour. Focused on the most fundamental building blocks, imaging genetics sets out to establish the connection between inter-individual genetic variation and observable differences in the brain (Bogdan et al., 2017). Historically, approaches in genetics have been focused on candidate genes, selected based on hypotheses of biological pathways tracing the

causal influence from gene to the trait of interest. This minimized the need for data, making genetics possible in the era predating advanced sequencing technology to yield important discoveries. However, it also has an important limitation: being largely driven by existing theory, it lacks the exploratory capacity to go far beyond what already exists in the body of knowledge. To alleviate this shortcoming, new approaches emerged in the mid-2000s, fueled by technological advances and the accumulation of vast amounts of data (W. Y. S. Wang et al., 2005). The most popular was the genome-wide association study (GWAS), a technique for detecting genetic variants associated with a trait through mass univariate testing (Uffelmann et al., 2021). These typically investigate variation in each atomic location in the genome, manifested through single nucleotide polymorphisms (SNPs), but can also be done for more complex sources of genetic variation such as copy number variations. Overall, the introduction of GWAS broadened the scope of genetic research, from narrow investigations of variants exerting a large influence on the phenotype at hand, to a broader search for variants contributing to complex phenotypes.

Over the last two decades, genetic studies, particularly those employing GWAS, have contributed to a wealth of discoveries in clinical neuroscience. These advances have been enabled by the collaboration and large datasets provided through large consortia, a prime example being the Enhancing Neuro Imaging Genetics through Meta Analysis (ENIGMA) consortium (P. M. Thompson et al., 2020). Using consortium data, or other techniques for pooling data from a wide variety of sources, modern GWAS commonly look for associations between genetic variants and clinical phenotypes based on data from hundreds of thousands of participants, enabling novel discoveries. A recent GWAS analyzing 7,500,000 SNPs in 320,404 individuals revealed 313 variants associated with SCZ (Trubetskoy et al., 2022). Similar efforts revealed 64 and 44 SNPs associated with BIP (Mullins et al., 2021) and MDD (Wray et al., 2018) respectively, in two GWAS based on 413,466 and 480,359 individuals. In AD, beyond the well-known impact of mutations in the apolipoprotein E (APOE) gene, the most recent GWAS identified 75 distinct genetic variants associated with AD risk, using full genome sequences from 788,989 individuals (Bellenguez et al., 2022). When combined into a composite score these yielded 1.9-fold increased risk for disease onset in the high-risk versus low-risk group. A recent meta-analytic approach including GWAS identified 551 putative genes that contributed to MS susceptibility. In PD, a meta-analysis across two GWAS with combined datasets of 429,225 individuals identified 41 significant associations

(D. Chang et al., 2017). These studies, and many more, have provided a broad foundation to investigate the genetic underpinnings of neuropsychiatric disorders.

4.3.1 Elucidating mechanistic and causal relationships via genetic associations

Although GWAS has played a major role in increasing the wealth of associations between genetic variation and neuropsychiatric disorders, it, like everything else, has its limitations. One major drawback is that as the approach rapidly identifies new associations, the magnitudes of their effects will be inversely proportional to their position in the sequence of discoveries. This could result in a never-ending stream of variants that contribute less and less to the phenotype at hand, while the total amount of variance explained remains modest (Goldstein, 2009). Relatedly, it has been pointed out that the expanding set of deleterious variants associated with a phenotype will contain a plethora of covarying but non-causal variants and does little to pinpoint the biological pathways through which they induce pathology (Tam et al., 2019). Consequently, they don't illuminate the underlying biological mechanisms, and as such have limited value both for advancing our knowledge of disorders and guiding development of treatments. To mitigate this, GWAS are often accompanied by subsequent analysis steps, to reduce the set of variants to those presumably causal and investigate the pathways through which they have an effect. The former is typically done via statistical fine-mapping, a statistical technique relying on the GWAS results, prior knowledge about the structure of the genome, and potentially empirical evidence about the role of the genes (Schaid et al., 2018) to nominate plausibly causal variants. Next, their impact is commonly elaborated through multi-omics, where measurements from cells or tissues can help pinpoint their role in the biological hierarchies of the body. In sum, these complex analysis pipelines provide a promising apparatus for understanding the biology underlying intricate imaging phenotypes.

A complementary method for increasing the value of identified genetic associations is to use them as instruments to understand causal interactions between high-level phenotypes themselves. This can be done with causal inference, in genetics operationalized through Mendelian Randomization (MR). MR is a method from genetic epidemiology (Davey Smith & Ebrahim, 2003), relying on the random distribution of genes from parents to offspring (Sanderson

et al., 2022), to draw causal conclusions based on associational data. This is implemented through an instrumental variable analysis, where the randomly distributed genetic variants are used as an instrument to investigate whether an exposure, such as a phenotypic risk factor, causally affects an outcome, for instance the onset of a disease (Emdin et al., 2017). To reach the lofty goal of causality, MR relies on three strict assumptions: that the genetic variant used as an instrument is associated with the exposure, that this association is not confounded by other variables, and that the instrument is only associated with the outcome through the exposure. The first is trivially verifiable through e.g. a GWAS, but the latter two are harder to confirm in empirical data. This is especially true in clinical neuroscience, where the correlation between disorders, complex traits, and intermediate phenotypes is high, making it hard to principally rule out confounding. Empirically, there has been shown broad horizontal pleiotropy for these phenotypes, indicating that they are associated with the same genetic variants (P. H. Lee et al., 2021), potentially through independent biological pathways. To alleviate this, a plethora of MR variants have been proposed, seeking to relax its fundamental assumptions. There is however still a responsibility on the independent researcher or analyst to ensure MR is a viable alternative in each specific use case (VanderWeele et al., 2014). Nonetheless, MR has provided, and will continue to be, a valuable tool for understanding the interrelation between complex traits, also in the neuroscientific domain.

4.4 Summary

One of the main aims of clinical neuroscience is to elucidate the relationship between biological variability in the brain and clinical phenotypes, including neuropsychiatric disorders. Structural MRIs allow us to investigate this relationship, by non-invasively providing information about the anatomy of individual brains. Subsequently, CNNs provide a promising technology to model these relationships based on imaging data to reach predictive goals. To achieve this, they will need to learn and leverage complex and potentially subtle patterns of neuroanatomical variation that could prove useful to describe the heterogeneity inherent to neuropsychiatric patients. However, the black-box nature of these models will make these potential insights incomprehensible. Thus, studies that rely on CNNs should be accompanied by efforts to understand them, through holistically relating their predictions and internal knowledge to other domains, and by applying techniques for explainability.

5 Research objectives

The overall aim of the PhD project was to utilize the representational capacity of deep learning models to learn complex, data-driven representations of the brain based on structural MRI data. These representations should adhere to two important principles: (1) they should be associated with clinically relevant variables, to provide information facilitating understanding and decision-making in the single-individual case, and (2) they should be understandable, to allow us to confidently use and learn from them. This overall goal was operationalized through different means in the three papers.

5.1 Paper I

In the first paper, we aimed to combine a state-of-the-art CNN architecture and a large dataset to leverage the predictive proficiency of deep learning for accurate and generalizable brain age predictions. Furthermore, we sought to understand whether brain age as learned by the model would coincide with a high-level notion of generalized brain health, through investigating associations with other measures. Finally, we wanted to explore whether variability in brain age, encoded in brain age deltas, and the internal representations learned by the model provided useful instruments for predicting neuropsychiatric disorders via transfer learning. Our main hypothesis was that the complex CNN trained on a large and heterogeneous dataset would robustly learn to predict brain age in a generalizable way, thereby summarizing overall brain health into a useful, reliable biomarker. Secondarily, we hypothesized that using the internal representations learned by the model would be more informative than the singular delta for predicting case-control status.

5.2 Paper II

The goal of the second paper was to further understand the underpinnings of brain age as learned by a deep neural network through applying a GWAS to brain age deltas akin to those in paper I. Next, we wanted to explore the significance of the identified genetic variants through in-depth investigations of their role in biological pathways, and their associations with neuropsychiatric disorders. Finally, we aimed to perform MR analyses enabled by these variants to investigate the causal relationships between brain ageing and the disorders.

Our main hypothesis was that genetic variants robustly associated with variation in the brain age delta would be associated with neuropsychiatric disorders and relate to a spectrum of biological processes in the body. Furthermore, we hypothesized that we would detect causal effects between brain age and the disorders previously associated with the biomarker.

5.3 Paper III

In the third paper, we focused on a single disorder, to investigate how XAI can be useful to describe heterogeneity within a single patient cohort. To reach this goal, we sought to train state-of-the-art CNNs in a composite clinical dataset containing patients with DEM and healthy controls. Around these models we planned to implement a pipeline including XAI to procure individual-level explanations of the model's predictions. We aimed to thoroughly evaluate the explanations to ensure their veracity. Finally, we wanted to see whether the explanations, encoding localization of pathology detected by the model, could be useful to support precise clinical decision-making in a cohort of MCI patients. Our main hypothesis was that the pipeline would procure explanations that were generally supported by current knowledge about where in the brain dementia manifests. Moreover, we hypothesized that their variability would reflect heterogeneity in the biological embodiment of the disorder, useful for supporting precision diagnostics for MCI patients.

6 Methodology

Paper	Images	Subjects	Age range (mean)	Females
I	56,095	56,095	3-95 (55)	52%
II	53,542	53,542	3-95 (55)	52%
III	20,306	2,913	45-97 (75)	45%

Table 1: Key characteristics of the complied datasets used in the three papers.

6.1 Data

An overview of the high-level characteristics of the datasets used in the three papers can be seen in Table 1. All of these were conglomerations of data coming from multiple sources, detailed overviews of which can be found in the independent papers. The data sources that were used beyond training brain age models and/or providing healthy controls for diagnostic analyses are described below and in Table 2.

- **AddNeuroMed (ANM)**: ANM is a cross-European study collecting imaging data and biological measurements to facilitate biomarker discovery for AD, acquired from six medical centers across Europe (Liu et al., 2011; Lovestone et al., 2007). We used structural MRIs for paper III, including patient groups with DEM and MCI that were defined by standardized thresholding of results from the mini-mental state examination (Tombaugh & McIntyre, 1992). The data was acquired through collaboration with Karolinska Institutet.

- **Alzheimer's Disease Neuroimaging Initiative (ADNI)**: ADNI is a publicly accessible dataset developed to facilitate early detection of AD and corresponding interventions, prevention strategies, and treatments. It contains brain scans from multiple imaging modalities, alongside biochemical measurements, behavioural data, and cognitive assessments from multiple centers in the US and Canada (Petersen et al., 2010; Weiner et al., 2012; Weiner et al., 2017). Structural MRIs were used for all three papers, genetic data was used in paper II, and cognitive assessments were

used in paper III. We used patients with DEM, diagnosed by ADNI investigators as probable AD using NINCDS/ADRDA criteria (Dubois et al., 2007), and MCI, diagnosed using a standardized clinical procedure (Petersen, 2004). The data was acquired through the online portal for the Image and Data Archive (IDA) at the Laboratory of Neuro Imaging (LONI).

- **Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL)**: AIBL is an Australian initiative mirroring ADNI, collecting neuroimaging and complementary data from two sites in Australia (Ellis et al., 2009; Fowler et al., 2021). Structural MRIs were used for papers I and III, including patients with DEM and MCI diagnosed with criteria equivalent to ADNI. The data was acquired through the IDA LONI portal.

- **Demgen**: Demgen is a subset of the Norwegian register of persons assessed for cognitive symptoms (NorCog), owned by Oslo University Hospital and administered by the Norwegian National Advisory Unit on Ageing and Health (Doan et al., 2017). Structural MRIs were used for papers I and III. This included patient cohorts with MCI and DEM, diagnosed according to NIA/AA 2011 criteria (Albert et al., 2011). The data was collected with approval from the Regional Committee for Medical and Health Research Ethics South-Eastern Norway (REK, application 2013/2283).

- **Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD)**: MIRIAD is a publicly available dataset created to investigate whether MRI provides a feasible outcome measure for clinical trials on the efficacy of treatments for Alzheimer's Disease (Malone et al., 2013). Structural MRIs were used for paper III, including patients with DEM diagnosed using NINCDS/ADRDA criteria for mild-moderate AD. The data was acquired from the XNAT platform.

- **Open Access Series of Imaging Studies 3 (OASIS3)**: OASIS3 is a publicly available dataset containing neuroimaging data, together with clinical and cognitive assessments and other measures from a broad demographic spectrum, compiled to facilitate research on cognition (LaMontagne et al., 2019). Structural MRIs were used for papers I and III. We included patients with DEM, including those with a diagnosis of either

probable AD, vascular dementia, or a combination, based on a standardized clinical assessment (Beekly et al., 2007; Morris et al., 2006). The data was acquired from the XNAT platform.

- **Oslo Multiple Sclerosis sample (OsloMS)**: Various neuroimaging modalities acquired from a collection of individuals scanned between 2012 and 2023 at the Oslo University Hospital through previous and ongoing research projects. Structural MRIs were used for paper I, including a patient cohort diagnosed with MS using McDonald 2017 criteria (A. J. Thompson et al., 2018). The data was collected and used with approval from REK (application 2016/102).

- **Thematically Organized Psychosis (TOP)**: The TOP study was initiated at the University of Oslo in 2002 and is an ongoing multicenter, multidisciplinary investigation of clinical, genetic, neuroimaging, pharmacological, and neurocognitive features of schizophrenia and bipolar disorders (Nesvåg et al., 2017). Structural MRIs were used for papers I and III, including patient groups with SCZ, BIP, and MDD diagnosed according to DSM-IV criteria (American Psychiatric Association, 1994). The data was collected and used with approval from REK (application 2009/2485).

- **UK Biobank (UKB)**: UKB is a large-scale population dataset collected from three sites in the UK, containing imaging data, genetics, and a wide variety of phenotypic information (Sudlow et al., 2015). Structural MRIs were used in papers I and II, alongside a broad array of phenotypic variables for paper I, and genetic information for paper II. No patient groups were included, but diagnostic information was utilized (based on ICD-10 codes, data field 41402) to exclude participants with mental (Chapter V) or neurological (Chapter VI) disorders from the brain age training set. The data was accessed under accession number 27412.

6.1.1 Imaging data

In all papers we applied a unifying preprocessing pipeline prior to modelling, modifying the MRIs through a six-step process:

1. Removal of non-brain tissue through skull stripping, using steps 1-5 of the recon-all pipeline from FreeSurfer 5.3 (Ségonne et al., 2004). Note that this also performs intensity normalization.

Dataset	I	II	III	Images	Subjects	Ages	Females	Diagnoses	
ANM			×	417	379	53-90	57%	DEM HC MCI	56 294 67
ADNI	×	×	×	20,912	2,498	50-97	44%	AD HC MCI	4,451 7,054 9,407
AIBL	×		×	905	588	58-96	53%	AD HC MCI	101 688 116
Demgen	×		×	277	277	38-89	44%	DEM MCI	134 143
MIRIAD			×	708	69	55-87	55%	DEM HC	465 243
OASIS3	×		×	3,039	1,098	42-95	57%	DEM HC	564 2,475
OsloMS	×			886	402	18-70	71%	MS	886
TOP	×		×	2,705	2,225	13-72	45%	BIP HC MDD SCZ	463 1,314 69 574
UKB	×	×		45,907	45,907	44-82	51%	HC	45,907

Table 2: Key characteristics of the data sources that was used for specific purposes across the three papers.

2. Transformation from .mgz to nifti format with mri_convert from FreeSurfer.

3. Reorientation to standard FSL space using fslreorient2std from the FMRIB Software Library (FSL) v6.

4. Linear registration to MNI152 space, performed by flirt (Jenkinson et al., 2002) in FSL with 6 degrees of freedom.

5. Removal of redundant voxels around the edges. This was implemented by extracting a central crop with bounds [6:173, 2:214, 0:160].

6. Rescaling intensity values to the range [0, 1] by division of the constant 255.

The resulting volumes had dimensions [167, 212, 160], and contained the greatest possible amount of brain tissue rigidly transformed to MNI152 space while minimizing the presence of extraneous tissues.

6.1.2 Genetic data

In paper II we utilized genetic data from UKB to perform a GWAS for the brain age predictions from our model. To collect this data, participants were split into subsets genotyped using two different, but similar, arrays (Wain et al., 2015), both resulting in ~800,000 genetic markers (Bycroft et al., 2018). Based on linkage disequilibrium (LD), the covariance of distinct, but related, genetic variants, the UKB team utilized these markers to impute specific variants per participant for ~100,000,000 SNPs. From these, we removed locations where the quality of the data was uncertain, and variants with very low frequency, to end up with a final sample of ~8,600,000 SNPs from 28,104 participants for analysis.

6.1.3 Phenotypic data

In paper I we performed a phenome-wide association analysis (PheWAS) to broadly explore properties that were associated with deviations in brain age. We utilized 402 diverse phenotypic variables from UKB which we qualitatively grouped into thirteen categories. Each variable was encoded according to PHESANT (Millard et al., 2018), a toolbox developed specifically for phenotypic data in UKB, alongside a standardized preprocessing pipeline.

In paper III we assessed the value of the relevance maps procured by our explainable pipeline by associating inter-individual variability in their spatial

content with performance on tests spanning various cognitive domains. These performances were recorded in 17 high-level summary scores extracted from 7 test batteries administered to 733 MCI patients in ADNI.

6.2 Modelling

6.2.1 Convolutional neural network architectures

The CNNs trained for all papers were variants of the Simple Fully Convolutional Network (SFCN), a brain age model tailored for three-dimensional structural MRI data, crowned winner of the Predictive Analytics Competition 2019 (PAC2019) brain age prediction competition (Gong et al., 2021; Peng et al., 2021). This model consists of a series of repeated convolutional blocks, each based on a convolutional layer and a max pooling layer, and as such resembles the models from the Visual Geometry Group (VGG) that broke major ground in improving image recognition of two-dimensional natural imagery in the early days of the CNN hegemony (Simonyan & Zisserman, 2015). Specifically, the backbone of SFCN consists of five blocks, each comprised of a 3x3x3 three-dimensional convolution, batch normalization, max pooling, and a rectified linear unit (ReLU) activation. These are followed by a single block that reduces the dimensionality of the feature space, implemented through a 1x1x1 convolution, batch normalization, and ReLU (Figure 7a). On top of this, the original model has a prediction head consisting of an average pooling layer, dropout, and a classification layer with a softmax activation, predicting an age probability distribution.

Based on this backbone, we experimented with different architectural variants suited for the exact problem at hand in the three papers, differing only in their prediction heads (Figure 7b) and the hyperparameter settings employed during training. In paper I we compared three variants constructed for predicting brain age, e.g. a single-valued, continuous output (potentially derived through further computational steps). First, we tested the unchanged, original, SFCN-model relying on soft classification, operationalized via a softmax activation in the final layer (SFCN-sm). Second was another model with multiple outputs denoting individual ages, but here interpreted through a ranking scheme, where each output predicted whether a participant was older than a given age (SFCN-rank). The ranking behaviour was implemented through sigmoid activations in the final layer. Third and last we employed a simple regression model

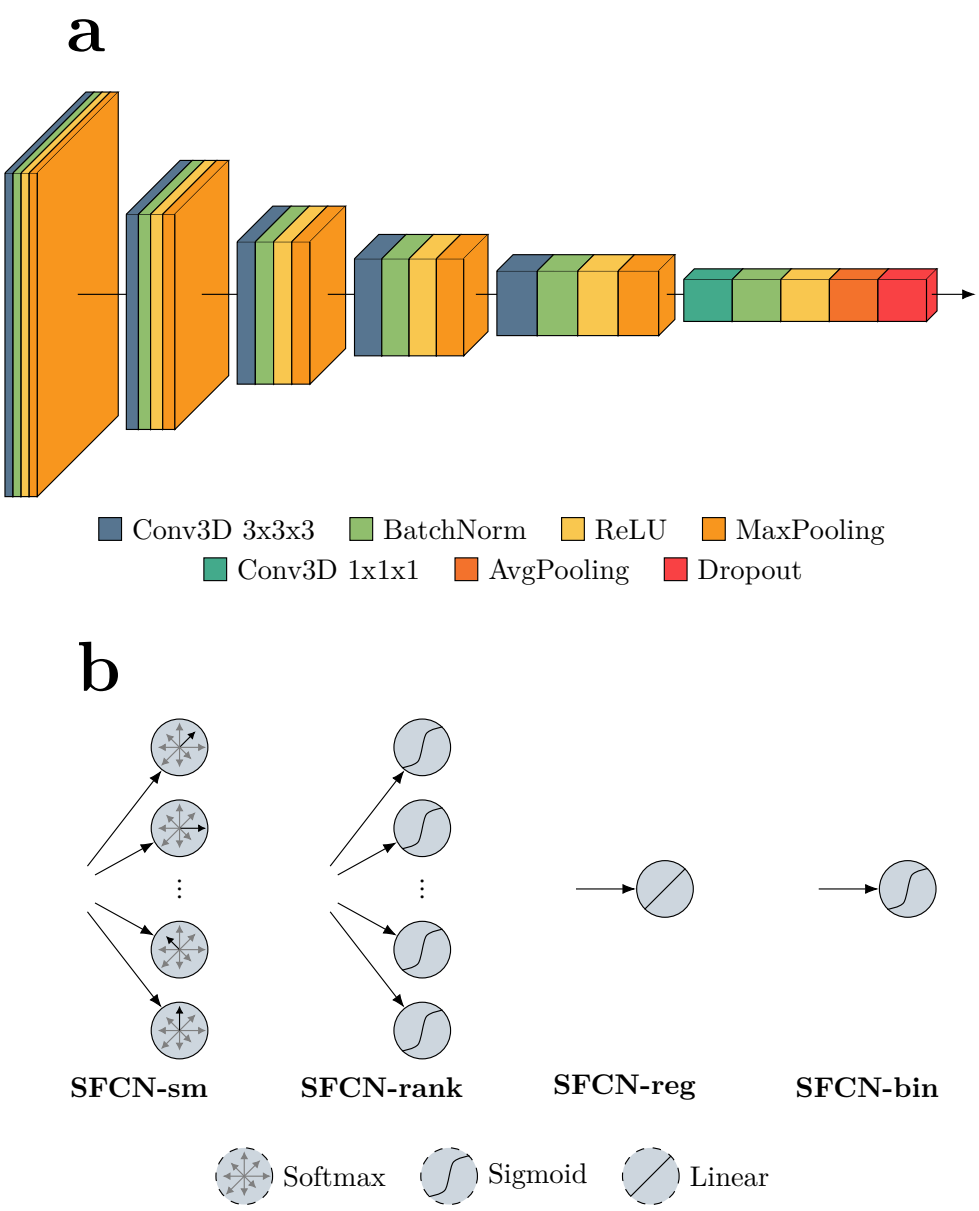


Figure 7: The convolution neural network architectures used throughout the thesis. (a) The Simple Fully Convolutional Network (SFCN)-backbone. (b) The prediction heads used for various predictive tasks across the three papers.

(SFCN-reg), with a single continuous output node without activation. In paper II we only used the SFCN-reg architecture, based on its superiority in paper I. In paper III we aimed to classify patients and controls and therefore implemented a variant with a binary classification head (SFCN-bin), comprised of a single output neuron with a sigmoid activation.

6.2.2 Model training and comparison

Like the architectures, the data splitting strategy employed varied between the papers depending on the use case. In paper I we combined a traditional train/validation/test approach (implemented through repeated 80%/20% splits) with an external dataset from unseen scanners to accurately assess model performance and generalization. In paper II we used a cross-validation approach without hyperparameter tuning, fitting a single model per fold (based on hyperparameters from paper I), to achieve out-of-sample predictions for all participants. In paper III we employed a nested cross-validation to facilitate both hyperparameter tuning and out-of-sample predictions. During tuning, we varied standard hyperparameters such as the learning rate schedule (including number of epochs), the dropout rate, and which augmentations to use. Each model was optimized to minimize a loss function tailored for its model architecture and specific task (Table 3). To select starting values and ranges for the different learning rate schedules we employed learning rate sweeps (L. N. Smith, 2017), where we evaluated how model performance in the training set changed as a function of the learning rate. In all three papers we selected the weights from the epoch yielding the best performance in the validation set, based on a metric suited for the predictive task (Table 3). All models were implemented in Python v3.8 for papers I and II, v3.9 for paper III, using Tensorflow (Abadi et al., 2015) v2.1 for papers I and II, and v2.6 for paper III, through the Keras interface (Chollet et al., 2015). The models were trained on Nvidia Tesla V100 and A100 GPUs.

6.2.3 Model evaluation

For model evaluation we employed the same two metrics as we did for epoch selection, tailored for each predictive task (Table 3). Brain age models (papers I and II) were evaluated by the mean absolute error (MAE), depending on a ground truth vector $y = [y_0, y_1, ..., y_{n-1}]$ and a vector of predictions $\hat{y} = [\hat{y}_0, \hat{y}_1, ..., \hat{y}_{n-1}]$. For each entry i, corresponding to an MRI scan, y_i

Model	Loss	Performance metric
SFCN-sm	Kullback-Leibler divergence	Mean absolute error
SFCN-rank	Mean binary cross-entropy	Mean absolute error
SFCN-reg	Mean squared error	Mean absolute error
SFCN-bin	Binary cross-entropy	Area under the receiver operating characteristic curve

Table 3: The losses and performance metrics used for each of the individual model architectures.

and $\hat{y}_i$ refer to the ground truth label (chronological age of the participant) and prediction (predicted brain age) respectively. MAE was then computed as $\text{MAE}(y, \hat{y}) = \sum_{i=0}^{n} |y_i - \hat{y}_i|$. For the dementia classification task in paper III, we evaluated models based on the area under the receiver operating curve (AUC). AUC also relies on two vectors y and $\hat{y}$, however here the entries y_i are binary and $\hat{y}_i$ continuous numbers in the range $[0, 1]$. In this scheme 1 typically encodes the positive class, in our case dementia patients, and the numbers $\hat{y}_i$ are interpreted as the probability of belonging to the positive class. Where most classification metrics rely on dichotomizing the predictions $\hat{y}_i$ to interpret them as either correct or incorrect, AUC instead considers the ordering of predictions between the two groups. Specifically, it evaluates whether the patients are predicted as having a higher probability of belonging to the positive class (e.g. predicted as patients) than the controls. This alleviates the need for calibrating the model through finding and setting an appropriate threshold for dichotomization, a process that should be performed based on domain knowledge. Furthermore, it alleviates the impact of class imbalance (Japkowicz, 2013). In sum, these properties render AUC generally more applicable in case-control applications than accuracy (Dinga et al., 2019), which remains the most widespread classification metric in the neuroimaging literature.

6.2.4 Operationalizing brain age

In papers I and II we trained CNNs to predict brain age. However, we weren't interested in the totality of ageing features detected by the model in any given brain, but rather the proportion of it that indicated a deviation from the normal ageing trajectory. This was isolated through the computation of a brain age

delta, referred to as a brain age gap (BAG) in paper II. To calculate the delta for each participant, we subtracted their chronological age from the predicted brain age, leaving a singular number that expresses their deviation from the normative curve, encoded in years. Due to properties of the statistical modelling processes that underlie brain age (Barnett et al., 2005), it is common to observe a linear bias of the predictions as a function of chronological age, forcing all predictions towards the mean age in the dataset. This bias would confound subsequent analyses by leaking information about chronological age through the delta, an issue that has normalized the use of correction procedures (de Lange & Cole, 2020) to ensure the delta is centered around zero across the entire age range. In our models we observed very little of this innate bias and chose to circumvent the correction procedure, known to have shortcomings of its own (Butler et al., 2021). Instead, we corrected for age in all subsequent analyses explicitly to minimize the potential of confounded results.

6.2.5 Explaining model predictions

In paper I we assessed what visual patterns were correlated with variation in the brain age delta, through two simple post hoc procedures. The first was a qualitative process where we visualized the average brains in groups with extreme deviations for multiple distinct age bins. The second was a slightly more advanced, quantitative, approach, where we correlated deltas with volumetric features from the images derived using FreeSurfer. These features were vertex-wise measures of cortical thickness and volumes of 45 brain regions (Fischl & Dale, 2000; Fischl et al., 2004), including six encoding hypointensities, defined according to the Aseg atlas (Fischl et al., 2002). For each of these measures, independently, we computed the Pearson correlation with brain age delta using numpy v1.20 (Harris et al., 2020). These correlations comprised a map that was overlayed a template brain and inspected visually.

In paper III we used LRP as a method for explaining the CNN trained for dementia classification. LRP is a post hoc explainability technique that we implemented on top of the trained models to form a pipeline procuring both predictions and relevance maps. For each input MRI given to the pipeline, the relevance maps represented a visual explanation of the prediction of the model, here corresponding to the probability assigned by the model that the MRI belonged to a patient with DEM. The full technical specification of the LRP implementation is described in the paper, and the implementation is available online.

Briefly, the relevance maps procured were three-dimensional volumes located in the same stereotactic space as the input, such that the relevance in each voxel denoted how much the corresponding input voxel contributed towards the prediction. To simplify the interpretations of the maps we employed parameterizations of the LRP process that upweighted positive relevance and removed negative relevance from the maps through a post-processing step. Consequently, the relevance maps were three-dimensional visual objects, spatially mirroring the input images, highlighting regions of the image containing evidence that was used by the model to arrive at the predicted probability of DEM.

6.3 Statistical analyses

6.3.1 Case-control analyses

In paper I we used diagnostic labels to probe the clinical utility of both the brain age deltas and the internal representations learned by the brain age model. Importantly, the latter were representations of the brain specifically related to age, as the model never saw diagnoses or any other cognitive or behavioural information during training. First, we assessed differences in brain age explicitly by contrasting the distribution of brain age deltas in six patient groups (AD, MCI, MOOD (BIP/MDD), MS, PSY (mixed psychotic disorders), and SCZ) to distributions from matched cohorts of healthy controls. This comparison was done using only data from scanners that were not seen by the model during training or validation. Next, to assess the same relationship in a predictive context, while also estimating the predictive gain offered by the learned representations, we constructed a simplistic transfer learning procedure to classify cases and controls in the same six cohorts. In this analysis, we employed 64 features from the second-to-last layer of the brain age model as predictors, alongside age, sex, and delta, and fit logistic regression models with l_1-regularization using scikit-learn v1.2.2 (Pedregosa et al., 2011). This was performed in a nested cross-validation loop, to achieve out-of-sample predictions while simultaneously tuning the regularization parameter λ. Finally, models were compared to assess the predictive value of various combinations of predictors, by concatenating out-of-sample predictions across all outer folds to compute a total AUC per predictor set. For each predictor, we assessed its information content by investigating whether adding it as a predictor increased the total AUC.

In paper III we used information from the explainable pipeline to fit a new

level of models using predictions and relevance maps for MRIs from a cohort of MCI patients. These models attempted to differentiate progressive MCI patients, e.g. patients that were diagnosed with dementia diagnosis at a subsequent timepoint, from a non-progressive group, comprised of both those that improved their condition at a later timepoint and those that remained stable throughout the data collection phase. Before these analyses, the high-dimensional relevance maps were decomposed into simplified feature vectors by a PCA across all MCI patients. The goal of this was to procure a manageable number of variables for subsequent modelling while retaining the most salient spatial information from the maps. In the subsequent analyses, the dimensions of the feature vectors, corresponding to the principal components of the PCA, were considered prototypical relevance maps, each encoding a stereotypical visual pattern observed in the MCI patients. Akin to paper I we approached the problem of prognostication via both a descriptive and a predictive route. First, we fit survival models where being given a diagnosis was treated as the terminal endpoint, using the feature vectors as predictors. We fit Cox Proportional Hazard models implemented in lifelines v0.27.1 (Davidson-Pilon, 2019), also including sex as a covariate for stratification. Performance of the model was evaluated in-sample, by assessing p-values to determine which components were associated with staying undiagnosed. To correct for multiple comparisons, we employed a p-value threshold determined via false discovery rate (FDR) correction (Benjamini & Hochberg, 1995). The outcome of this analysis was a narrowed-down set of prototypical relevance maps that were significantly associated with a future diagnosis, allowing for a qualitative interpretation of how the localization of early dementia-related pathology related to progression. Next, we reformulated the question of prognosis in a predictive context, simulating a clinically realistic prognostic scenario. Here, we fit binary classifiers to predict which of the MCI patients would progress into dementia at several fixed timepoints in the future. These models were logistic regression models using different combinations of age, sex, probability of dementia at the current timepoint (predicted by the original CNN), and feature vectors as predictors (representing relevance maps), fit with an l_1-regularization term using scikit-learn. As before, this was performed in a nested cross-validation loop, to achieve out-of-sample predictions while tuning λ. Model performance was assessed using the AUC, this time calculated independently for each outer fold. Having multiple assessments of performance for each model allowed us to more rigorously test which predictors yielded a significant improvement, by doing pairwise comparisons of the distribution of

paired AUCs in a one-sided Wilcoxon signed rank test implemented in scipy v1.6.3 (Virtanen et al., 2020).

6.3.2 Genetic analyses

In paper II we investigated the genetic foundations of differences in brain age, operationalized through the brain age delta based on brain age predictions from CNNs. These models were fit in a 5-fold cross-validation to achieve out-of-sample deltas for all participants. We first performed a GWAS meta-analysis to identify SNPs associated with variations in the delta, implemented using PLINK version 1.9 (C. C. Chang et al., 2015), by running independent GWAS for each of the five folds. The results from these five independent analyses were unified using PLINK's inverse variance-weighted meta-analysis to produce a single effect size and significance level per SNP (C. H. Lee et al., 2016). Based on the resulting genome-wide significant associations, we first resolved the issue of multiple SNPs representing the same haplotypic variant by clumping the results through FUMA (Watanabe et al., 2017), resulting in a single lead SNP per genomic region. Each of these was linked to adjacent genes through the Ensembl Variant Effect Predictor (VEP) tool (McLaren et al., 2016). Next, we identified plausible causal variants within each region using FINEMAP (Benner et al., 2016) and assessed whether these affected the expression levels of the genes via lookups in the GTEx v8 eQTL portal (the GTEx Consortium, 2015).

Based on the results from the preceding analyses we looked for overlapping associations with five disorders, based on summary statistics from recent GWAS for AD (Jansen et al., 2019), BIP (Mullins et al., 2021), MDD (Wray et al., 2018), PD (Nalls et al., 2019), and SCZ (Trubetskoy et al., 2022). First, we calculated the amount of variance explained by genetic variability for each disorder, commonly referred to as SNP-based heritability, and the correlation between disorder and delta using LD score regression (Bulik-Sullivan, Loh, et al., 2015) implemented in ldsc (Bulik-Sullivan, Finucane, et al., 2015). We also assessed polygenic enrichment, indicating whether the genetic variants associated with one trait are also associated with another, through conditional quantile-quantile (QQ) plots (Chen et al., 2017; Lo et al., 2017), conditioning on both brain age and the disorders. Here, enrichment was assessed visually by identifying whether the variants significantly associated with the conditioned trait had p-values lower than expected by random chance in the target trait, using increasingly stricter thresholds of significance for inclusion.

In addition to the correlational analyses above, we employed multiple MR models to disentangle the causal relationship between the brain age delta and the five disorders. These were performed bi-directionally, both using brain age as a risk factor and each of the disorders as an outcome and using the disorders as risk factors and brain age as an outcome. In each analysis, we selected genetic variants as instrumental variables based on the GWAS for the corresponding risk factor and assessed the ratio of effect sizes between the risk factor and outcome for each instrument. When these ratios are similar, meaning that the genetic profile that affects the exposure and outcome is similar, and the fundamental assumptions underlying MR are met, this allows for the interpretation that the risk factor causally affects the outcome. However, evaluating the validity of these assumptions was challenging, as they often are in real-life data, so we employed five different variants of MR to alleviate them in various ways, in addition to the conventional inverse-variance weighted model (Relton & Davey Smith, 2012). The weighted median (wMed) model relaxes the assumption that the genetic variants used as instruments need to be associated with the exposure (Bowden et al., 2016). Egger regression (Bowden et al., 2015) tests and controls for horizontal pleiotropy, when genetic variants affect both the risk factor and outcome independently, through introducing an intercept in the statistical model. The pleiotropy residual sum and outlier (MR-PRESSO) model identifies outlier variants caused by horizontal pleiotropy that inflate the causal estimate (Verbanck et al., 2018). The robust adjusted profile score (RAPS) model applies robust regression techniques to minimize the impact of instruments with pleiotropic effects (Zhao et al., 2020). Finally, the causal analysis using summary effect (CAUSE) model accounts for pleiotropic effects both when they affect the risk factor and the outcome through the same biological mechanism, and when they are uncorrelated (Morrison et al., 2020). To unify results across these six variants we used a majority voting scheme, interpreting only results where four or more of the methods agreed as reliable.

6.3.3 Validating relevance maps

To ensure the classifier in paper III had learned to recognize patterns plausibly associated with DEM, and that these were conveyed through our LRP implementation, we performed two sanity checks. First, we averaged the relevance maps across all correctly predicted patients to form a singular relevance map representing the average patient and compared it with a statistical reference en-

coding preexisting knowledge of the localization of dementia-related pathology from the scientific literature. The statistical reference map was generated with an activation likelihood estimation implemented in GingerALE v3.0.2 (S. B. Eickhoff et al., 2012; S. B. Eickhoff et al., 2009; Turkeltaub et al., 2012). The basis for the map was 394 experiments from 124 publications related broadly to DEM, collected with Sleuth v3.0.4 (Laird et al., 2005). We non-linearly registered the average relevance map, such that both it and the statistical reference map were three-dimensional volumes in MNI152-space, both containing continuous values. To perform the comparison, we binarized both maps at different percentiles, and computed the Dice-Sorensen coefficient, denoting the fraction of overlap, between the voxels that surpassed the binarization threshold on each side. This produced a curve, tracing the overlap as a function of the percentile used for binarization, which was examined qualitatively. To determine whether the overlap surpassed chance levels, the process was repeated for three other relevance map-generating pipelines, where the procured relevance maps had no association with DEM-related pathology.

The second sanity check was quantitative, centered around the CNN that was trained to predict probability of DEM. Here, we used the relevance map produced for each image to iteratively perturb it, for each iteration occluding the region that was identified as contributing the most towards the prediction. Additionally, for each iteration, we ran the occluded image through the classifier, recording how the perturbation influenced its prediction. This allowed us to plot a curve, tracing the prediction as a function of the number of perturbations. Subsequently, we calculated an area over the perturbation curve (Samek et al., 2017), quantifying how quickly the CNN's ability to recognize patients faded as important regions were sequentially being removed. Again, we compared the curve produced by the DEM-pipeline with the three alternative pipelines, to see whether it quantitatively was better than chance at detecting regions important for classifying dementia.

6.3.4 Additional analyses

Although the imaging models, case-control classifiers, and genetic analyses described so far cover most of the analytical components employed throughout the three papers, we supplemented them with various other analyses to further understand and exploit what was learned by our CNNs. In paper I we performed the PheWAS to associate inter-individual variability in brain age deltas

with variation across 402 other phenotypic variables. These associations were assessed univariately, using a linear model to model delta as a function of each phenotypic variable, while controlling for age and sex. The models were implemented in statsmodels v0.13.2 (Seabold & Perktold, 2010). Significance was assessed by assessing the p-value denoting the contribution of the phenotypic variable in each independent model. We accounted for multiple corrections controlling the FDR.

In paper III we associated visual patterns in the relevance maps from the dementia classifier with performance on a neuropsychological test battery, to assess whether the location of dementia-related pathology detected by the model was associated with clinical symptom load. Here, we also used the simplified representations of the relevance maps derived from the PCA described above, namely feature vectors with dimensions representing prototypical relevance maps. Each of these prototypes was associated univariately with 17 summary scores from the test battery via linear models in statsmodels. In addition to the feature vectors we corrected for age, sex, and dementia prediction from the original CNN, to isolate the effect of the localization of pathology, as opposed to the total pathological load. As above, significance was assessed by evaluating the p-values for the individual prototypical relevance maps in their respective models, with an FDR-corrected threshold.

6.4 Ethical considerations

The work in this thesis was performed partially using data acquired locally, regulated by multiple approvals from REK (applications 2009/2485, 2013/2283, 2016/102, 2019/943). Data management and privacy issues concerning those studies have been evaluated and approved by the University of Oslo (UiO) and/or Oslo University Hospital. Data from various online sources and collaborators have also been utilized, governed through a variety of data sharing agreements. All analytical work was performed on UiO's services for sensitive data (TSD), a platform facilitating storage and analysis of sensitive data in a secure environment. The research was performed in compliance with Norwegian law for conducting medical and health research (the Health Research Act).

Beyond what is explicitly governed by laws and regulations, there will always be ethical aspects that should be considered when doing research. This holds true also for this thesis, and I will elaborate some of them in the following paragraphs. First, I will discuss some relevant ethical challenges that arise

when doing medical research with diagnostic and prognostic aims, particularly using advanced predictive technology. Next, I will elaborate on issues that arise when collecting and using data for research purposes. Finally, I will delve into some of the ethical considerations pertaining to the use of AI, first when using it as a tool to derive scientific knowledge, and secondly pertaining its broader role in society.

We generally want people to be healthy and content, an objective frequently operationalized through medical research and practice. When someone is not healthy, we seek to treat them, ideally returning them to a state of well-being. To recognize which remedy is appropriate a diagnosis is decreed, effectively clumping the present patient with other patients that suffer from the same, or at least a very similar, disease. This allows the prescription of a treatment, based either on a mechanistic understanding of the nature of the illness, or empirical evidence of what has successfully treated the disease historically. This general approach has formed the backbone of evidence-based medicine in the last half-century. And while there have always been challenges with how and when a diagnosis should be given, in this framework this has generally been a question of balancing risks and benefits to maximize the probability of a good outcome. However, the complexity of deciding when a diagnosis should be given increases substantially as the definition of a disorder obfuscates. Furthermore, the potential upside of giving a diagnosis is reduced when this knowledge may provide a limited basis for deciding upon effective treatment, a reality for many of the complex disorders addressed in this thesis (Durães et al., 2018; Leichsenring et al., 2022). Oppositely, the potential downside increases dramatically if the disorder is associated with social stigma, such that simply giving the diagnosis has a negative effect on the patients, another aspect that has been described in these disorders (Hinshaw & Stier, 2008). Combined, these factors indicate that these diagnoses should be prescribed with care. Similar arguments can be made against prognosis: it is complicated to determine whether the insight that one may be at risk to contract a life-changing, potentially untreatable, neuropsychiatric disorder is beneficial, especially if that knowledge does little or nothing to alleviate the risk. Through the development of predictive technology, this thesis can be reasonably said to contribute towards automatizing these diagnostic and prognostic processes, prompting contemplation on whether this is morally and ethically defensible.

To counteract the problems outlined above, I will point out reasons to be optimistic about the use of AI and technology to diagnose neuropsychiatric dis-

orders. Most importantly, early, precise, and personalized diagnosis has proven effective for slowing disease progression and reducing symptom load for a range of disorders, even when curative treatment is not currently available (Rasmussen & Langerman, 2019). The importance of this point is further emphasized by findings revealing that individuals are more willing to listen to recommendations when these are based on individualized risk as opposed to general advice (Carver et al., 2022). If this can be achieved through individualized predictive models, it seems more appropriate to ask how they can be implemented safely, rather than categorically denouncing them. Regarding the morality of labelling individuals with diagnoses when this doesn't entail treatment, there are positive sides that can counteract negative stigma. Specifically, being given a diagnosis can provide patients with a new lens through which they can understand themselves and their challenges (Werkhoven et al., 2022), potentially reducing the burden of a condition. However, neither of these arguments should make out the cornerstone of a moral defense for predicting disorders and outcomes. Instead, I will argue that prediction can provide a useful tool in our quest to understand the complex mechanisms of these enigmatic conditions. First, I will assertively state that I consider this pursuit of the causes of disease a moral imperative. This might seem redundant, but there are plausible arguments for how a more careful dissection of individuals along behavioural axes of variability can have negative consequences, such as overdiagnosis (Moynihan et al., 2012), overmedicalization (Angell, 2011a), and a general constriction of normality (Angell, 2011b). While definitively important issues these are massively outweighed by the tremendous negative effects inflicted by the conditions considered in this thesis. However, it is worth mentioning the unreasonability of considering these diagnoses as a single entity in this regard, given their innate differences. Nonetheless, given the intricacies of demarcating them, and the general focus on methodology and technology here, it is a necessary simplification.

Prediction, when judiciously applied, can play a pivotal role in the exploratory process of understanding a phenomenon. To accentuate this view, I will return to the representation learning perspective embraced in the introduction. While predictions from a statistical learning model are commonly seen through the lens of statistics, representing probabilities or likelihoods, it is equally justifiable to interpret them as information-theoretical quantities. Here, they simply represent a quantum of information, a compressed version of the data that was input to the model that retains some properties while discarding others, according to specific rules. In cognitive and clinical neuroscience, it can

be argued that there is a lack of intermediate levels of description and explanation between the biological and psychological domains (Rolls, 2021). One of the main concerns of the field should be the construction of a conceptual hierarchy that bridges this gap. However, reconciling these domains will require navigating a complicated intermediary landscape of vast combinatorial possibilities. Here, predictive models, through their capacity for processing and compressing information, can potentially play the role of landmarks in this exploratory pursuit. This is best depicted in an idealized imagined setting with a predictive model able to combine information to arbitrary degrees of complexity, e.g. a DNN (Hornik et al., 1989), and infinite data. Here, the concordance between prediction and labels will be indicative of the information overlap between the input and output space of the model. In clinical neuroscience, this could allow for quantitative reasoning about what can possibly be derived from neuroimaging data, and its interrelation with clinical constructs. From this perspective, the predictive models take on an inductive role in the quest for knowledge, where the focus is not on the practical utility of their predictions per se, but rather on their proficiency as information processing entities. To summarize, having predictive capabilities does not necessarily imply using them for single-subject predictions. While I believe that also the latter is morally defensible in some cases, this disentangles the two processes, simplifying the moral argument for the former.

The scientific process has always relied on data from the real world to refute or strengthen hypotheses. When the subject matter under investigation relates to personal life, this data will necessarily be highly sensitive. The absolute dependence on sensitive and personal data has fostered the development of both concrete judicial structures and ethical guidelines for the management and usage of such data for research purposes. In the age of big data, data mining, and deep learning, some of the central pillars of these governing frameworks are being challenged. One such is the principle of data minimization, which states that a researcher, or any other data-collecting entity, should collect the minimal amount of sensitive data sufficient to realize a purpose (The Norwegian Personal Data Act, §5.1c). This stands in contrast with the exploratory nature of modern ML approaches, where data is not used simply to test hypotheses, but also to generate them, implying that the ideal starting point is a dataset containing as much information as possible. This desire for breadth also potentially contradicts the related principle that data should be collected towards a specific goal (The Norwegian Personal Data Act, §5.1b). Consequently, it is crucial to weigh

the potentially detrimental effects it could have upon participants in studies to effectively relax these restrictions through reformulations of the intended purpose and goal, against the potential benefits of having access to the data. As this thesis relies solely on data already collected, this issue is not of direct relevance. Nonetheless, it is important to keep in mind when using external data, to ensure it was collected in accordance with high ethical standards and that the current research falls within the collected consent. For the data used here, this is the case. A related issue emerges when data from various sources are combined to provide the broadest possible basis for explorative modelling. This can potentially incite situations where data that was originally, independently, not sufficient to identify individuals, but through combination this guarantee is lost. Although this thesis combines a lot of data from different sources, this never entailed augmenting data from the same individual, and thus this concern is alleviated.

If the data we used was collected with appropriate consent and according to high ethical standards, stored securely, and not shared, it sounds reasonable to assume that privacy concerns are mitigated. However, there are other potential issues to be aware of. One such that arises particularly in research developing predictive models, and maybe specifically DNNs, is the public sharing of trained models, a practice inherited from the broader open-source community. In a sense this practice seems unproblematic: the shared models are large matrices of numbers that are unintelligible to humans. However, these numbers are derived from the data that was used to train them, in this context highly sensitive data governed by strict rules and regularization. For generative models, a different class of models than what we used, recent research has shown it possible to extract samples that were used for training after the model was deployed (Carlini et al., 2023). While the exact technical route for reverse engineering the types of models we used is less obvious, this still opens the possibility that we are somehow leaking sensitive data through our shared models. In addition to privacy concerns, there are general problems regarding governance and ownership of models distilled from personal, sensitive data collected under specific consents. Taken together, these issues problematize the sharing of pretrained models. However, there are also moral arguments favouring the other side. Most importantly, this is a practice that enhances the capacity of the collective scientific community to develop better predictive models and put them to use (Touvron et al., 2023), in addition to fostering reproducibility (European Commision et al., 2020). For the models trained in this thesis the latter weighed

heavier, and they are open-sourced under non-commercial licenses.

Another ethical concern that has always existed in research, but is exacerbated through the sharing of pretrained models, is the representativeness of the data that is used. Traditionally, various biases have challenged the generalizability of empirical findings. In medical research, a systemic error that has persisted through the years is gender bias. Historically, most studies have included mostly male participants, leading to knowledge, and subsequent development of interventions, that do not generalize to females (Holdcroft, 2007). However, this has improved in later years, and while the effects of gender differences are still not always properly accounted for, many of the large neuroimaging datasets are relatively balanced with respect to the sexes of their participants. Conversely, there are still severe selection biases in ongoing data collection efforts concerning geography, demography, and race (Ricard et al., 2023). This limits the generalizability of findings and could lead to differences in the healthcare solutions available to different populations and groups. Overcoming this requires changes in data acquisition practices, but until then it is important to be specific about the limitations this entails. The exacerbation concerning shared models comes from the fact that this lack of generalizability will now be implemented practically in tools that are made available to others. Furthermore, it might be hard for users to recognize this problem, and, even when it is known, difficult to determine when a new population is so far out-of-distribution that the model is no longer reliable. Ideally, this should be handled by extensive testing of where the model behaves as intended by its developers. However, as this is not a practically feasible solution, it is at the very least important for practitioners to be open and explicit about what data was used to train their models.

It is important to note that many of the issues outlined above are magnified when AI models initially developed for scientific endeavours are translated into publicly accessible software, for instance in clinical decision support systems. Now, unrepresentative samples used for training models does not solely lead to invalid inferences, but potentially social injustice or medical malpractice. And poor data management can lead to personal and sensitive data collected with consent to pursue scientific questions ultimately increasing shareholder value in commercial companies. The effort to regulate modern AI is just getting started with the EU AI Act, a proposed law for the regulation of AI systems in Europe to be implemented in early 2024. Although this development is of broad interest to everyone working in the field, there is one topic specifically relevant to this thesis that incites reflection. If commercialized, the technology developed here would

presumably be categorized as medical devices, putting them in the high-risk group of products (AI Act, Recital 30). One of the demands for products in this group is transparency, an aspect that is operationalized in this thesis through explainability. In the legislation, transparency is defined as "[...] enabling users to understand and use the system appropriately" (AI Act, Article 13.1) and that "Users should be able to interpret the system output and use it appropriately" (AI Act, Recital 47). This reasserts the central role of the user in XAI outlined earlier. And practically, it entails that empirical, quantitative investigations of the explanations of AI models, such as the ones performed here, must be supplemented by trials with potential users of the system to understand whether they are useful. Ensuring that explainability truly fosters understanding will be imperative to develop and deploy AI that conforms to ethical and judicial expectations moving forward.

I will end this chapter by adding my voice to a collective concern raised in response to the rapid development of AI in the last year. It is common practice in many scientific branches that academics and commercial entities collaborate to advance the research frontier. In AI, this dynamic appears to be somewhat shifted towards commerce, with many breakthroughs in the last decade originating from large technology companies (Brown et al., 2020; Jumper et al., 2021; Vaswani et al., 2017). This can be explained by the immense commercial potential of the technology fostering private investments, potentially accelerating its progression for the overall good. For many years this happened in a tolerable fashion, where private actors contributed towards advancing the state-of-the-art, necessarily promoting their commercial interests, but while making theory, and often even code and models, accessible through standard academic channels. This changed in 2020, when OpenAI, a research company created to facilitate the safe development of artificial general intelligence (OpenAI, 2015), decided not to open source their groundbreaking large language model GPT-3 (OpenAI, 2020). The rationale given was the model's potential to cause societal harm, particularly its proficiency at generating misinformation (Brown et al., 2020). Instead, the technology was exclusively licensed to Microsoft (Scott, 2020), about a year after OpenAI changed from a non-profit to an investor-friendly "capped profit" structure to meet their rising capital needs (OpenAI, 2019). In 2022 ChatGPT was released (OpenAI, 2022), a complete end-user application for interacting with the newest, still proprietary, GPT models, followed by an optional, paid, premium service (OpenAI, 2023). The debate on the ethical aspects of open-sourcing advanced deep learning models is still ongoing

and is a topic too complex to describe in detail here. However, considering the preceding paragraphs urging general caution when developing AI methodology in scientific contexts to avoid overstepping ethical bounds, it is worth contemplating that there are actors in the field with access to more data and resources, striving for plausibly less noble goals than the scientific pursuit of knowledge, operating under less stringent ethical frameworks. The proficiency of tools developed by private actors could lead to situations that necessitate their use, even if they were developed with unsatisfactory ethical standards. Finding a balance between regulating and empowering different actors to ensure progress while mitigating the risk of harm must be a priority for both regulators and practitioners of the field in the years to come.

7 Summary of papers

7.1 Paper I: Deep neural networks learn general and clinically relevant representations of the ageing brain

Background: Over the last decade, brain age has emerged as an intuitive and sensitive marker of brain health. However, disparities in data acquisition have proven troublesome for the statistical methodologies underlying brain age models, raising questions about the generality of what they learn, and hindering clinical adoption. Facilitated by the accumulation of vast datasets and theoretical advances, deep learning provides opportunities for refining brain age and enhancing its potential for clinical implementations. Empirical studies based on large, heterogeneous datasets are imperative to investigate whether they can overcome differences in data acquisition, scrutinize their clinical utility, and further understand what the measure encodes.

Methods: We compiled a large, heterogeneous dataset ($n = 53,542$) of structural MRIs to fit brain age models using variants of a state-of-the-art convolutional neural network. To assess generalizability, we performed out-of-sample testing with scanners unseen by the model. Next, we tested for associations between variability in predicted brain age, structural brain measures, and a variety of phenotypes. Finally, we employed the brain age model in a transfer learning setting to predict neuropsychiatric diagnoses (AD, MCI, MOOD, MS, PSY, SCZ) to assess its clinical relevance.

Results: The best model achieved state-of-the-art generalization, with an MAE of 3.9 years in data from unseen scanners. Increased brain age was associated with widespread cortical thinning and reduction of subcortical volumes. Furthermore, we found correlations with a multitude of phenotypes, including diabetes, cardiovascular measurements, smoking, and alcohol consumption. Utilizing the representations learned by brain age model for transfer learning yielded fair predictive power for detecting patients with MS, AD, and MCI, (AUCs > 0.7), and modest differentiation for SCZ, PSY, and MOOD (AUCs ≈ 0.6)

Conclusion: A CNN trained to predict brain age based on large and heterogeneous data can alleviate common issues with generalization, and its predictions coincide with measures known to be associated with overall and brain-specific health. Furthermore, both brain age and the learned representations underlying it are predictive of neuropsychiatric disorders.

7.2 Paper II: Genetic architecture of brain age and its causal relations with brain and mental disorders

Background: BAG, the difference between the apparent age of the brain and the chronological age of an individual, has been linked to a multitude of neuropsychiatric disorders. However, the genetic underpinnings of BAG have not received much attention, even though they could elucidate the role of brain ageing in neuropsychiatric disorders. Beyond genetic associations, MR offers a methodology to probe the causal relationships between BAG and these disorders, potentially explicating its value as a clinical biomarker.

Methods: We used a state-of-the-art CNN to predict brain age in a subset of participants from the UK Biobank ($n = 28,104$) from which genetic data was also accessible. Based on these predictions we calculated BAG and performed a GWAS to identify genetic variants that were associated with it. Next, we applied a battery of methods to refine the genetic signal and identify potential biological pathways related to BAG. Finally, we investigated the relationship between BAG and five neuropsychiatric disorders (AD, BIP, MDD, PD, SCZ) through polygenic overlaps and bi-directional MR analyses.

Results: The brain age models were accurate, achieving consistent MAEs of < 2.5 years in unseen data. The GWAS revealed eight independent loci significantly associated with variability in BAG, seven of which were novel. The supporting analyses indicated that these genetic variants influenced 54 genes, which had intricate expression patterns in the brain specifically, and other tissues across the body. The MR analyses indicated that higher BAG was protective against Parkinson's disease, whereas increased genetic risk for AD and BIP were causally related to higher brain age. However, qualitative inspections revealed that these causal relationships should be interpreted with care.

Conclusion: Our study represents the largest GWAS on brain age to date, identifying seven novel genetic variants contributing to its variation. These variants were associated with genes expressed in a range of tissues across the body, emphasizing the sensitivity of brain age as a composite marker of generalized health. The causal relationships between BAG and neuropsychiatric disorders remain elusive.

7.3 Paper III: Characterizing personalized neuropathology in dementia and mild cognitive impairment with explainable artificial intelligence

Background: Dementia, afflicting more than 55 million individuals worldwide, is a neuropsychiatric condition that incurs a monumental detrimental effect on society. Dementia patients are a heterogeneous group, both in terms of pathology and clinical manifestation. Early detection and personalized characterization of the disease would facilitate accurate prognosis, widen the window for early interventions and potentially alleviate uncertainty about the future of individual patients. XAI has the potential to accurately localize dementia-related pathology to facilitate personalization and provide a translation technology for dementia diagnosis and prognosis.

Methods: We trained CNNs to differentiate patients with DEM ($n = 854$) from healthy controls ($n = 854$) using structural MRIs and implemented LRP on top of the model to form an explainable pipeline, procuring both individual predictions and relevance maps to explain them. We validated the relevance maps extensively, first through a comparison with existing knowledge of pathology recorded in the literature, and secondly by assessing the predictive value of the information they encoded. Finally, we employed the pipeline in a cohort of MCI patients ($n = 1256$) to investigate its potential for prognosis and personalization of the diagnosis at an early stage.

Results: The best model achieved satisfactory discrimination performance, reaching an out-of-sample AUC of 0.9. The first validation revealed substantial overlap between the average relevance map from our pipeline and the reference map produced through a meta-analysis of the literature. The second validation affirmed the importance of the regions detected in the relevance maps for predicting dementia in a quantitative setting. Information from the pipeline in the MCI cohort enabled the ability to differentiate progressive from stable MCI patients with an AUC of 0.9 after 5 years. Differential localization of pathology encoded in the relevance maps showed associations with impairments in distinct cognitive domains.

Conclusion: XAI presents a promising translational technology to support clinical decision-making in the diagnosis and prognosis of dementia. The outputs of our explainable pipeline was both predictive of disease progression and associated with individualized clinical manifestation of the disease, both essential capabilities to enable precision medicine in dementia in the years to come.

8 Discussion

Through this thesis, I have explored the hypothesis that DNNs are capable of learning representations of the brain from neuroimaging data that can help elucidate the biological variability that underpins neuropsychiatric conditions. Moreover, I have demonstrated that these representations are associated with quantifiable biological and physiological processes in the body and the brain, to underscore that they are not simply visual patterns, but rather represent composite and informative biological measures. Finally, I have related the representations with existing diagnostic labels, with a focus on neuropsychiatric disorders with a behavioural or cognitive aspect, to highlight their clinical potential. In the upcoming section I will continue discussing the results from the papers, to highlight independent aspects that warrant attention, elucidate their interconnections, and place them into the broader context provided by contemporaneous research in the intersection of AI and neuroscience.

8.1 Deep learning-derived brain age as a robust marker of generalized brain health

In papers I and II we trained CNNs to predict brain age based on a large and heterogeneous sample, to contribute to a broad literature emphasizing the usefulness of brain age as a marker of generalized brain health. The advantages of using complex deep learning models to predict brain age as opposed to simpler models, such as the relevance vector machine underlying the original brainAGE framework (Franke et al., 2010), are not necessarily obvious. The most evident benefit of transitioning to more expressive models is generally the increase in predictive performance, in brain age exemplified by the SFCN winning PAC2019 (Gong et al., 2021; Peng et al., 2021). However, more accurate brain age models are not trivially more useful. Recently, there have been put forth arguments favouring both accurate models (Hahn et al., 2021) and models that allow for a moderate amount of variation in their predictions (Bashyam et al., 2020; Bashyam et al., 2021) even though the latter necessitates worse predictive performance as measured by standard metrics. The proposed advantage of loose-fitting models is that variability in their predictions reflects biological variation, rendering them more sensitive towards pathological alterations manifest in various patient groups. This is epitomized in the argument that a perfect brain age model, predicting the correct chronological age every

time to yield brain age deltas of exactly zero, is useless, as these will not covary with any phenotype of interest. While this is tautologically true, it is also worth further contemplation. Although the outputs of the model would be neither interesting nor useful, it can be argued that the model itself necessarily would be. Having learned to navigate the complexities of neuroanatomy to perfectly and invariably recognize age seems to imply a profound understanding of the changing brain, knowledge that appears invaluable.

To examine the opposite side, the main argument against loose-fitting models is the practical improbability of determining what causes their worse fit. While it can be caused by biological variability, it can also arise from modelling errors or imaging artifacts, and it is implausible that these two can be told apart for an individual prediction. Following this I advocate the view that we should pursue more accurate models, under two important presuppositions. First, that better fits are not achieved by actively reducing heterogeneity in the dataset, in paper I ensured via the large dataset and external validations. And second, venues for exploiting the knowledge encoded in the models beyond their raw predictions must be explored, exemplified through our transfer learning scheme where we used the internal representations of the model to predict diagnostic status.

Potentially more devastating than uncertainties about modelling best practices are controversies regarding the true nature of brain age, such as the discussion on whether it is sensitive towards ongoing rates of aging, or if it rather captures differences that are relatively stable across the lifespan. Empirical evidence has been presented both in favour of the former (Franke & Gaser, 2012; Høgestøl et al., 2019) and the latter (Vidal-Pineiro et al., 2021). Related is the problematization of the relationship between age-related patterns encoded in these brain age predictions and neuropsychiatric pathology (Cole & Franke, 2017), and specifically to what degree the former encompasses the latter (Heinrichs, 2023). This is not specific to brain age, but rather part of a broader philosophical discussion concerning the interrelation between age-related and pathological processes (Hayflick, 2004; Holliday, 2004). A third matter that has been raised is the limited interpretability of brain age due to its composite nature (Cole & Franke, 2017). In my opinion, the last objection falls on its own accord: it is exactly the ability to aggregate and consolidate information that makes brain age valuable. This does not mean it's not worthwhile to inquire what underlies brain age. Both paper I and other studies (S. M. Smith et al., 2020) have indicated that there are levels of abstraction beneath the singular brain age that can be useful and interesting, and this direction should be ex-

plored further. As a broader stance against the issues presented here, I concur with the pragmatic view (Cole & Franke, 2017), pointing towards the practical utility of brain age. The PheWAS and diagnostic predictions in paper I add to a wealth of empirical evidence amassed for the sensitivity of brain age predictions towards alterations in the brain related to biological processes. This view should not impede investigations into the ontological foundation of brain age, but neither should ontological uncertainty hinder practical work toward exploiting its utility. While the amount of evidence for the usefulness of brain age predictions continues to mount, there are, to my best knowledge, no implementations of brain age models tailored for clinical use available today. To achieve this should be a priority for practitioners in the field moving forward.

8.2 Elucidating brain age via genetic associations

Another promising approach for utilizing brain age as learned by our CNNs is as an intermediate phenotype between low-level biology and high-level cognition and behaviour, including clinical diagnoses. Paper II adds to a growing body of knowledge about the genetic foundations of brain age (Jonsson et al., 2019; Ning et al., 2020; S. M. Smith et al., 2020) and its genetic associations with neuropsychiatric disorders (Kaufmann et al., 2019), by both replicating earlier findings and identifying novel genetic associations. It also gives further credence to the notion of the brain age delta as a heritable trait (Cole & Franke, 2017; Jonsson et al., 2019; Kaufmann et al., 2019). Beyond the variants themselves, an intriguing finding was the extensive scope of biological pathways associated with them. This implicates brain age as a measure sensitive towards an array of biological processes. This finding can be seen in relation to more general findings indicating a broad genetic foundation underpinning other biological clocks (Rando & Chang, 2012), potentially linking brain age with a more comprehensive notion of biological aging. However, this inference is complicated by an earlier study finding no association between brain age and methylation age (Cole et al., 2018), signifying that the two measures are sensitive to complementary information. The latter hypothesis has been strengthened by recent reports that an age prediction model combining brain and body measures predicted age more accurately than a standalone brain age model (Beck et al., 2023). In sum, this outlines an intricate relationship that should be elucidated through further efforts. However, it also highlights the potential of brain age in combination with other biological ages to provide a comprehensive array of abstract, high-level

biological measures that can holistically support precise clinical decisions.

A surprising result from paper II was the lack of overlap we observed between the genetic variants underlying brain age and those underlying neuropsychiatric disorders, contradicting earlier findings (Kaufmann et al., 2019). There are multiple plausible explanations for this discrepancy, beyond differences in methodology. Under the assumption that the genetic architecture is complex and polygenic, and the individual effects are small, the datasets used by us and others for brain age GWAS are still relatively small (Nishino et al., 2018). This is an assumption that has proven true for other complex neuroimaging phenotypes (Roelfs et al., 2023). This presumed polygenicity could also influence the validity of the MR analyses we performed. However, if this was the case, and some of the genetic variants affected brain age and one of the disorders through horizontal pleiotropy, this should yield inflated causal results and potentially false positives (Bowden et al., 2015). Instead, we saw little reliable evidence for causal relations overall. Like above, a potential explanation for this is the limited sample size of the GWAS, although imaging genetics studies based on simpler imaging phenotypes have revealed evidence for causal relationships based on similar sample sizes (Guo et al., 2022). It is possible that the complexity of brain age as a measure, specifically when derived from a DNN, conflates the genetic variation underlying it, giving it a broad pleiotropic foundation. If so, it could be more meaningful to look for genetic signals at levels of abstraction underlying the singular, composite, measure (S. M. Smith et al., 2020), such as the internal representations we used for transfer learning in paper I.

8.3 Characterizing heterogeneity with explainable artificial intelligence

In paper III we assessed brain heterogeneity through XAI by leveraging the localization of pathology enabled by our explainable pipeline built for classifying patients with DEM. However, to do this confidently, we first validated the veracity of the explanations through two validation procedures. Extensive validation of these approaches is crucial, given the general shortcomings of these methods that have been repeatedly shown (Adebayo et al., 2020; Kindermans et al., 2019; Sixt et al., 2020). These shortcomings are potentially exacerbated in neuropsychiatric disorders where even the ground truth diagnostic label is subject to heterogeneity and subjectivity (Martin et al., 2023). On top of this, the importance of concretizing the added value of applying XAI in individual

studies has been emphasized (Lipton, 2017), incurring a need to establish why specific validation procedures are applicable in given situations. In our first validation, we observed that the regions where the model detected evidence for DEM corroborated knowledge existing in the literature, similar to what has been reported by others (Böhle et al., 2019; Dyrba et al., 2021; D. Wang et al., 2023). This validation approach reliant on preexisting knowledge is important to assess the fidelity of the explanations, and it represents a first step towards evaluating whether the model utilizes information that appears plausible to a human expert, facilitating trust (Tonekaboni et al., 2019). The second validation was focused on assessing whether the regions that were detected as informative by our pipeline had an impact on the predictions of the model, and as such was concerned with internal consistency as opposed to external validity (Samek et al., 2017). Taken together, these two validations provided confidence that our pipeline procured reliable information related to DEM, enabling subsequent analyses.

It is interesting to contemplate the requirement for AI used in clinical situations to be explainable, and the necessity of these explanations to be thoroughly validated. Currently, decisions taken in these situations are not supported by algorithms but rather made by humans. It has been shown that explanations provided by humans to justify their decisions do not necessarily reflect the actual decision processes underlying them (Holzinger et al., 2019; Johansson et al., 2006). Furthermore, it has been reported that humans tend to overestimate their ability to understand the decision processes of others, purely based on the assumption that they resemble their own (Bonezzi et al., 2022). Beyond individual humans, it has been argued that medical practice, both historically and contemporaneously, is mostly predictive, not explanatory (London, 2019). Building upon this argument leads to the claim that it is not intrinsic, virtuous qualities of potential explanations that enable trust in the decisions made by clinical personnel, but rather confidence in their expertise, accumulated experience, and proven track record. It is possible to take the same stance towards AI (Cappelen & Dever, 2021) and weigh its predictive capacity above and beyond its ability to explain itself. Nonetheless, this is not what upcoming legislation aims for. One of the reasons underlying the skepticism towards decisions made by DNNs is their propensity to take shortcuts, giving the appearance of having understood and properly modelled a phenomenon, when in reality they have found a clever way of surpassing it altogether (Geirhos et al., 2020). This can make them fail in ways that appear incomprehensible, that would never occur

in the hands of a human expert. Providing a mechanism for detecting these situations is one of the goals XAI is striving for. However, it is critical to ensure the new methodology does not simply provide a new level of shortcuts leveraging our propensity for seeking meaningful explanations (Ghassemi et al., 2021; Lipton, 2017). This underscores the importance of multifaceted validation procedures.

Beyond the extensive validation, the main innovation in paper III was to utilize individualized explanations for personalized diagnostics and prognostics. This is a possibility that has been mentioned in the literature (Martin et al., 2023) and employed in other domains (Jin et al., 2021), but, to the author's best knowledge, has not been explored with regards to neuroimaging data. The technicalities underlying this possibility deserve elucidation. In general, ML approaches can be broadly categorized as either supervised or unsupervised learning (Goodfellow et al., 2016). The former aims to train models that solve problems where the answer is known, operationalized through labelled samples used for training. These are the approaches that produce run-of-the-mill classifiers that differentiate cats and dogs, or cases and controls based on diagnostic labels. Although this can be knowledge-producing (Plis et al., 2014), the main benefit of supervised approaches is to systematize and operationalize existing knowledge through automatization. Conversely, unsupervised learning is exploratory in nature, encompassing approaches where models find patterns in data to optimize a mathematical objective not linked to a specific label. A prime example is clustering, where samples are grouped based on their innate characteristics, rather than predefined groups. The ability of unsupervised learning to provide new characterizations of data can be useful, mostly when these covary with interesting dimensions of variability, detecting using additional data. However, this is not necessarily the case, and the model can learn perspectives that do not appear useful (Altman & Krzywinski, 2017).

In between these two extremes is semi-supervised learning, where labelled data is used to guide the model, but not completely determine its output (Davatzikos, 2019). In paper III we showed that something similar can be achieved through a combination of supervised learning and XAI. Here, the diagnostic labels guided the process by providing a foundation for training the model, but the pipeline in its totality produced relevance maps with a spatial richness far exceeding what existed in the original data. Furthermore, we were able to link the spatial information in the relevance maps with clinically useful measures. This highlights the ability of the approach to perform precise characterization

to support personalized clinical decision-making beyond what is possible today.

Beyond its usefulness in practical, clinical situations, the pipeline in paper III exemplifies a broader possibility to use XAI to produce new scientific knowledge. In complex neuropsychiatric disorders, accurate CNNs combined with techniques like LRP could help reveal parts of the brain containing aberrations in patient groups. This is made possible even when these consist of patterns that are intricate and subtle, and, importantly, without preexisting labels describing them. However, this endeavour should be preceded by scrutinizing whether this allegedly explainable technology constitutes a valid epistemic tool for deriving scientific knowledge. In the case where a predictive model approximates a relation representing an underlying phenomenon, there are two processes that can be explained: The decision process underlying the predictions of the model, and the process governing the phenomenon itself (Srećković et al., 2022). Where the latter is the goal of scientific inquiry, it is the former can be approximated by XAI. This description highlights two points of interest. First, the explanations provided by XAI are two approximations above the underlying phenomena, plausibly causing them to be imprecise (Shmueli, 2010). Second, current predictive models, including the ones used in this thesis, are associational in nature, relying on statistical associations in data to procure predictions. This means they should not be trusted to accurately depict the causal mechanisms that science commonly seeks to uncover. Furthermore, there has been a lot of emphasis on the recipient of the explanation in XAI, to guarantee that the explanation fosters subjective understanding. While this is valuable from a practical standpoint, it also implies that the explanation should not necessarily be taken as depicting an objective fact (Páez, 2019). Taken together, these do not invalidate the use of XAI as an epistemic tool but plausibly limit its utility: it appears more appropriate as an exploratory tool to form hypotheses that can be investigated further through other means, as opposed to playing a confirmatory role.

8.4 The current state of deep learning in neuroimaging

Throughout this thesis I have argued principally for the particular aptness of deep learning to discover complex, non-linear patterns of neuroanatomical aberrations that potentially underlie neuropsychiatric disorders based on neuroimaging data. This argument has primarily been based on the capacity of DNNs to abstract and learn hierarchical representations (LeCun et al., 2015), properties

seemingly appropriate for dealing with the intricacies distinctive to the hierarchical organization of the brain (Rolls, 2021). If this argument holds one could imagine that we now, a decade into the era of deep learning, would encounter a neuroimaging literature brimming with empirical evidence supporting the preeminence of deep learning. But while there are sporadic successes following the introduction of DNNs, such as the SFCN winning PAC2019, neither this has caused leaps in our ability to predict behavioural phenotypes. Instead, studies have reported that deep learning methods reach approximately the same predictive efficacy as traditional ML for predicting high-level phenotypes (He et al., 2020; Schulz et al., 2020), undermining the theoretical argument of its superiority. This discrepancy deserves attention.

First and foremost, some of these comparative studies have been criticized for not enabling DNNs to reach their full predictive potential (Abrol et al., 2021). Furthermore, the two-step process including preprocessing underlying the traditional analyses has been seen to foster suboptimal modelling practices, enticing a form of double-dipping that could induce a bias in their results (Arbabshirani et al., 2017). There have also been more general reports of suboptimal validation practices (Whelan & Garavan, 2014) and overfitting, issues empirically backed by the recurrent finding that predictive performances reported in neuroimaging studies seem to decrease as a function of sample size (Wolfers et al., 2015). While there are no guarantees for better validations when using deep learning, there has been an increased focus on the importance of the topic in recent years (Scheinost et al., 2019), plausibly improving practices. However, it has also been argued in the opposite direction, claiming that overfitting has worsened with the introduction of deep learning models (Davatzikos, 2019). In the end, comparing performance across studies is hard, to the level where they have been called meaningless (Arbabshirani et al., 2017), and it is plausibly more fruitful to focus on best practices going forward. Throughout all three papers in this thesis, we have pursued using as rigorous validations as possible: in paper I we employed an external test set, providing the highest level of credence in the model's ability to generalize. In paper II, where we relied on predictions for all participants to enable the GWAS, we performed a cross-prediction scheme to ensure out-of-sample predictions for the subsequent analyses. In paper III, where a smaller dataset allowed us to train more models, we employed a similar, nested, scheme, to achieve out-of-sample predictions while tuning hyperparameters. Overall, these rigorous validations have allowed us to trust the results achieved by our models, to use them to generate knowledge through subsequent

analyses.

It is worth noting that the SFCN architecture used extensively here, a relatively simple VGG-like architecture, outperformed other approaches, including more advanced CNN variants, in a presumably objective comparison in PAC2019. This could indicate that it is still early days in the adoption of deep learning methods in the realm of neuroimaging and that there are still ample benefits to reap. Importantly, it seems plausible that the key to enabling the adaptation of a cascade of increasingly predictive models is not the time needed to translate them to the new domain, but rather the availability of data. The combination of all openly accessible neuroimaging datasets is still orders of magnitude smaller than ImageNet (Madan, 2022). As pointed out before, this is an issue especially evident for clinical data (Venkatraghavan et al., 2023). Thus, there are purely quantitative reasons that could explain why deep learning is still to reach its full potential in clinical neuroscience. With the ongoing amassment of data, coupled with solutions (Plis et al., 2016) and repositories (S. B. Eickhoff et al., 2016) for data sharing, and decentralized training (Rootes-Murdy et al., 2022), is it plausible that this challenge can be overcome in the years to come.

On the topic of ImageNet, it is worth emphasizing the tremendous impact the recurrent, objective assessment of the state-of-the-art via ILSVRC had on the collective capacity for image recognition based on natural images. A similar pattern has been observed in neuroimaging (Arbabshirani et al., 2017), through PAC2019 (Fisch et al., 2021), the ADHD-200 competition (Milham et al., 2012), TADPOLE (Marinescu et al., 2019), and others (Bron et al., 2022). Competitive modelling has been historically intertwined with a tradition of sharing state-of-the-art models, also a major contributor towards improving deep learning models and facilitating their implementation into practical solutions. In this regard, we have contributed by making the brain age model from paper I publicly accessible, which has allowed it to be scrutinized by others (Dörfel et al., 2023; Hanson et al., 2023). Furthermore, it has been used to elucidate the ontology of brain age (Holm et al., 2023), its clinical potential (Persson et al., 2023), and its technical limitations (Korbmacher, Wang, et al., 2023). Overall, there are reasons to be hopeful that the collective efforts of researchers, pulled together via, among other things, competitions and the sharing and dissemination of data, models, and software, will enable the continued maturation of deep learning in neuroimaging, to generate insights and translational technological solutions.

8.5 Limitations and methodological considerations

The work in this thesis is subject to multiple limitations, the most important of which will be described here. The first is regarding the generality of the terminology that is used. The term neuropsychiatric disorders is used broadly to describe a range of disorders. In reality, the empirical work performed in the thesis concerns itself with a set of distinct diagnoses, simplifying reality dramatically. This is not to undermine the existence of others or their complexities, nor is it making value-based judgments upon which deserve to be the object of scientific study. Instead, it is the result of practical necessity, limited by time and the availability of data. And while I believe that the methodology developed and tested is more broadly applicable than what I have been able to show, it is also worth being aware of the dangers of overgeneralization. As such, empirical studies are necessary before conclusions are drawn concerning the usefulness of the methodological frameworks outlined here towards other conditions. A similar statement can be made with about the imaging modalities used in the present thesis. Again, I talk broadly about the aberrations that are detectable in the brain via neuroimaging data and methods for modelling them, whereas all three papers rely solely on structural T1-weighted imaging data. Once again, this represents a practical necessity, and there is arguably a multitude of other imaging modalities that could have been used instead. Furthermore, these have shown to complement each other (Arbabshirani et al., 2017), also with respect to the exact problems and methodological approaches discussed in this thesis (Rokicki et al., 2021; S. M. Smith et al., 2020; Teipel et al., 2015). As such, further studies are required to investigate the proficiency of the methods outlined here towards other modalities, ideally in combination, to assess whether they are more broadly useful.

8.5.1 Intricacies of clinical validations

While we have tried to perform validations to ensure the clinical validity of the models and the representations they learn in our papers, these have been validations relying on clinical aspects of the available data, namely diagnostic labels, not validation in clinical contexts. The importance of this difference has recently been emphasized in the literature (Varoquaux & Cheplygina, 2022). Thus, despite our promising results, it is important to highlight some outstanding issues that complicate clinical translations of the methodologies outlined here. First and foremost is the authenticity of the clinical scenarios we have attempted to

emulate. For instance, in paper I, we validate the utility of the brain age model through a case-control comparison, an approach that has been subject to broad critique which I have outlined in this thesis. A more appropriate validation of clinical utility would be to employ the model to support a realistic clinical decision. One example of this would be differentially diagnosing patients from a clinical cohort, a direction we have later explored (Persson et al., 2023). In paper III we plausibly come closer to validating our pipeline in a realistic clinical use case, when we validate the applicability of the relevance maps for prognosis. However, ideally, these validations should happen in clinical settings, with clinical personnel, based on actual patients and the decisions that arise concerning their care. The key to releasing the translation potential in the technologies explored here relies on continued transdisciplinary, collaborative efforts (Herzog, 2022). We have provided the empirical foundation to facilitate these efforts, and further research should focus on exploiting this towards genuine clinical value.

In addition to this overarching issue comes technical complications. First is the reliability of the models. While none of the papers in this thesis has explicitly investigated the extent of this, later applications of the brain age model from paper I have revealed that predictions for a single individual vary with years based on scans acquired days apart (Korbmacher, Wang, et al., 2023). This is concerning: while the model appears reliable in research contexts relying on large groups (Dörfel et al., 2023), this greatly diminishes the utility of its predictions to support personalized, precise, clinical decisions. Overcoming this issue relies on further developments focused specifically on increasing robustness, and, importantly, detecting situations where the model is plausibly wrong, for instance via accompanying the predictions with uncertainty measures (Hahn et al., 2022). Another technical limitation of the models we have developed is their reliance on high-quality, research-grade MRI scans. Thus, our models can't be deployed directly in clinical settings, where the MRI scanning protocols are typically different. Approaches have been proposed to alleviate this problem, such as training models directly on clinical-grade data (Wood et al., 2022), or using software to enhance clinical images to research-grade quality (Iglesias et al., 2023) prior to modelling (Valdes-Hernandez et al., 2023). However, in the case of our already trained models, it seems more plausible to consider finetuning based on clinical data. It has been shown that CNNs can adapt to new scanners using transfer learning (Karani et al., 2018), also in small datasets (R. Wang et al., 2023). This direction contributes towards exploiting the technology and models developed here towards clinical translations and should continue to

receive attention from researchers.

8.5.2 The limits imposed by existing samples

Although we strove to make the datasets for all three papers as large and heterogeneous as possible through substantial efforts in collecting and consolidating data, the generalizability of our findings relies directly on the characteristics of the data that was used, as is always the case in empirical research (Woo et al., 2017). However, in the case of this thesis the concern is doubled, as the goal was not only to generate and disseminate knowledge but also to train models useful to others. Thus, it is worth describing what those characteristics were, to minimize the potential for downstream errors. For papers I and II, UKB made up the bulk of the compiled dataset. UKB is a population-based dataset from the UK containing participants aged 44-83 years, and it is plausible that our results are misleading or biased outside that cohort. We have seen this effect in practice, observing qualitatively that the performance of the model was notably worse outside this age range. To alleviate the potential detrimental effects somewhat we tested our model in an external test set, to better assess generalization. While this is above and beyond what is commonly done in the field (Woo et al., 2017), it is worth mentioning that also the participants in our testing set were mostly of white European backgrounds, and thus inferences beyond this group should be made with care.

Similarly, for paper III, most of the data came from ADNI, indicating that the both model and the empirical results derived from it are most probably descriptive of a northern American cohort. Here, being a clinical sample, it is also worth including that the participants were selected via the application of strict exclusion criteria (Petersen et al., 2010), limiting the generalization of findings further. In addition to the specifics of samples that are given by practical constraints, such as their geographical location, or explicit aims of the study, necessitating a specific age range or clinical cohort, it is worth contemplating how various selection biases can affect our results in various adverse ways. It is hard to isolate and quantify exactly what these effects are, and common to instead urge caution in the interpretation of results. However, I believe that we observed at least one instance of selection bias that allows us to infer that they occur in the results presented here: in paper I, we observed that being born outside the UK was associated with a substantially younger-looking brain. This was identified based on a relatively small group of what are plausibly highly

selected individuals, namely those that both move to another country and participate in a scientific study. It is unnecessary to speculate further upon what truly underlies this association, but it underscores the importance of careful interpretations of the results from broad associational studies based on observational data such as the three papers presented here. To alleviate these concerns, future data acquisition should include a broader spectrum of parcitipants.

8.5.3 The validity of cross-sectional inferences

Another point worth contemplating is the types of inference made possible from cross-sectional data. Although paper III uses longitudinal data implicitly to train the predictive models and explicitly to explore the prognostic capacity of our pipeline, our studies and models are essentially cross-sectional. Brains vary cross-sectionally, but also in terms of trajectories of change over time (Fjell & Walhovd, 2010), and the interrelations between the two and their individual and combined impact on behaviour and cognition is largely unknown. It has been shown that longitudinal study designs are far more sensitive for detecting group differences (Steen et al., 2007). However, due to their increased demands concerning cost and time, they are commonly disfavoured compared to cross-sectional designs. One of the downsides of using cross-sectional, observational, data, is the limitations this causes in terms of drawing causal conclusions, as it is impossible to determine what preceds what in time. This is however also true for many longitudinal designs. The statistical approaches we use, both deep learning and others, excluding MR, are essentially associational in nature. All in all, the consequence is once again an urge for caution. Some measures can be taken to minimize the potential for erroneous interpretations of conclusions drawn in studies based on cross-sectional data (Kraemer et al., 2000). One of these is the precision of the terminology being used, where temporal terms indicating change over time should be replaced with static terms denoting cross-sectional differences. In the context of this thesis, this is particularly relevant in the case of brain age, for instance in the use of "accelerated", that have plausibly contributed to disagreements about what the measure truly encompasses. To avoid these issues, we have avoided temporal language in papers I and II. However, to remove any potential doubt, it is worth stating explicitly that this thesis has focused on cross-sectional data and modelling with the limitations that entail, and should be interpreted as such. To elucidate the interrelations between our cross-sectional results and longitudinal change, further studies are

needed based on other data sources and study designs.

8.5.4 Architectural choices

There are reasons to believe DNNs are innately appropriate for modelling brains. After all, the former were inspired by the latter. However, this does not necessarily mean that CNNs are equally appropriate to model information in brain scans. Like all other predictive models, CNNs are built with a set of inductive biases, assumptions about the problems they are meant to solve explicitly built into the architecture to maximize their performance. One example shared with general DNNs is the assumption that the data contains hierarchal information, presumably appropriate for brain scans. Two other inductive biases in CNNs specifically are locality and spatial invariance. The former assumes that information close in space is probably more associated than that which is far apart. The latter that visual patterns can occur anywhere in an image and still mean the same. Neither of these seem apt for brain imaging data. There are other DNN architectures that avoid these assumptions, and I will describe two for which I have particularly high hopes. The first is vision transformers (ViTs), a class of architectures based on the popular transformer architectures for language processing (Vaswani et al., 2017), adapted for image data (Dosovitskiy et al., 2021). These can combine features from standard CNNs with advanced attention mechanisms to flexibly vary the assumption of locality across scales. This would allow for constructions that assume adjacent voxels are interconnected up to some point, after which longer distance connections are allowed, seemingly even more suitable for neuroimaging data than what is offered by CNNs. The second is graph neural networks (GNNs), which surpass the assumption that the input data is organized in space completely and rather see it as nodes in an unstructured graph (Kipf & Welling, 2017), another assumption that arguably suits brain imaging data better than those underlying CNNs.

There are historical reasons why this thesis has focused on the SFCN architecture, namely that the work was undertaken right after PAC2019. However, there are also more principled arguments for why simpler CNN architectures still are reasonable choices for modelling neuroimaging data currently. Most notably, standard ViTs have proven to be even more data-hungry than CNNs (Liu et al., 2021). And GNNs require preprocessing prior to modelling to turn the MRI data into graphs, potentially reintroducing the problems with human interference described earlier. Nonetheless, innovative approaches have shown that

both are becoming feasible alternatives for modelling based on structural MRI data (Bessadok et al., 2022; Bi et al., 2023). Through continued development of theory and growing datasets, their adoption in the neuroimaging community and subsequent impact will continue to grow.

9 Concluding remarks

The main aim of this thesis was to exploit the expressive power of deep learning models to learn complex patterns in neuroimaging data to quantify biological heterogeneity and elucidate its role in neuropsychiatric disorders. All three papers involved training state-of-the-art CNNs for various predictive tasks based on large, heterogeneous datasets, and then utilizing the knowledge encoded in the model for subsequent analyses. We have explored a variety of techniques to contextualize and understand both the predictions and the underlying representations of the models and demonstrated how they can support clinical decision-making. Importantly, this has been done using strict validations to ensure that our findings are reproducible. Taken together, the work in this thesis exemplifies that deep neural networks can learn interesting representations of the brain and that these can be both meaningful and useful.

Future studies should continue to develop expressive models supported by the growing availability of neuroimaging data, and employ innovative approaches to understand and use the representations they learn. However, it is equally important that they focus on the translational potential innate in this technology, to advance it towards clinical implementations. This relies on collaborations between researchers and clinical experts, to ensure that the technology that is developed can be used to solve realistic problems in real-life clinical scenarios.

References

Aarsland, D., Påhlhagen, S., Ballard, C. G., Ehrt, U., & Svenningsson, P. (2012). Depression in parkinson disease—epidemiology, mechanisms and management. *Nature Reviews Neurology*, *8*(1), 35–47. https://doi.org/10.1038/nrneurol.2011.189

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2015). Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 19.

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., & Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature Communications*, *12*(1), 353. https://doi.org/10.1038/s41467-020-20655-6

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2020). Sanity checks for saliency maps. *arXiv:1810.03292 [cs, stat]*. http://arxiv.org/abs/1810.03292

Aisen, P. S., Cummings, J., Jack, C. R., Morris, J. C., Sperling, R., Frölich, L., Jones, R. W., Dowsett, S. A., Matthews, B. R., Raskin, J., Scheltens, P., & Dubois, B. (2017). On the path to 2025: Understanding the alzheimer's disease continuum. *Alzheimer's Research & Therapy*, *9*(1), 60. https://doi.org/10.1186/s13195-017-0283-5

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Altman, N., & Krzywinski, M. (2017). Clustering. *Nature Methods*, *14*(6), 545–546. https://doi.org/10.1038/nmeth.4299

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders : DSM-IV*.

Angell, M. (2011a). The epidemic of mental illness: Why. *The New York Review of Books*, *58*(11), 20–22.

Angell, M. (2011b). The illusions of psychiatry. *The New York review of books*, *58*(12), 20–22.

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, *145*, 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry–the methods. *NeuroImage*, *11*(6 Pt 1), 805–821. https://doi.org/10.1006/nimg.2000.0582

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE*, *10*(7), e0130140. https://doi.org/10.1371/journal.pone.0130140

Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C., & Mechelli, A. (2021). Machine learning for brain age prediction: Introduction to methods and clinical applications. *eBioMedicine*, *72*, 103600. https://doi.org/10.1016/j.ebiom.2021.103600

Balestriero, R., & baraniuk. (2018). A spline theory of deep learning. *Proceedings of the 35th International Conference on Machine Learning*, 374–383. https://proceedings.mlr.press/v80/balestriero18b.html

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bashyam, V. M., Erus, G., Doshi, J., Habes, M., Nasrallah, I. M., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Satterthwaite, T. D., … CARDIA studies, t. P. A. d. C. (2020). Mri signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals world-

wide. *Brain*, *143*(7), 2312–2324. https://doi.org/10.1093/brain/awaa160

Bashyam, V. M., Shou, H., & Davatzikos, C. (2021). Reply: From 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. *Brain*, *144*(3), e32–e32. https://doi.org/10.1093/brain/awaa455

Beck, D., De Lange, A.-M. G., Gurholt, T. P., Voldsbekk, I., Subramaniapillai, S., Schindler, L., Hindley, G., Leonardsen, E. H., Rahman, Z., Van Der Meer, D., Korbmacher, M., Linge, J., Leinhard, O. D., Kalleberg, K. T., Engvig, A., Sønderby, I., Andreassen, O. A., & Westlye, L. T. (2023, November 3). *Dissecting unique and common variance across body and brain health indicators using age prediction* (tech. rep.). Radiology; Imaging. http://medrxiv.org/lookup/doi/10.1101/2023.11.03.23298017

Beck, D., de Lange, A.-M. G., Pedersen, M. L., Alnaes, D., Maximov, I. I., Voldsbekk, I., Richard, G., Sanders, A.-M., Ulrichsen, K. M., Dørum, E. S., Kolskår, K. K., Høgestøl, E. A., Steen, N. E., Djurovic, S., Andreassen, O. A., Nordvik, J. E., Kaufmann, T., & Westlye, L. T. (2021). Cardiometabolic risk factors associated with brain age and accelerate brain ageing. *Human Brain Mapping*. https://doi.org/10.1002/hbm.25680

Beekly, D. L., Ramos, E. M., Lee, W. W., Deitrich, W. D., Jacka, M. E., Wu, J., Hubbard, J. L., Koepsell, T. D., Morris, J. C., Kukull, W. A., & Centers, N. A. D. (2007). The national alzheimer's coordinating center (nacc) database: The uniform data set. *Alzheimer Disease and Associated Disorders*, *21*(3), 249–258. https://doi.org/10.1097/WAD.0b013e318142774e

Beheshti, I., Mishra, S., Sone, D., Khanna, P., & Matsuda, H. (2019). T1-weighted mri-driven brain age estimation in alzheimer's disease and parkinson's disease. *Aging and Disease*, *11*(3), 618–628. https://doi.org/10.14336/AD.2019.0617

Bellenguez, C., Küçükali, F., Jansen, I. E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A. C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., Holmans, P. A., Boland, A., Damotte, V., van der Lee, S. J., Costa, M. R., Kuulasmaa, T., Yang, Q., de Rojas, I., Bis, J. C., … Lambert, J.-C. (2022). New insights into the genetic etiology of alzheimer's disease and related dementias. *Nature Genetics*, *54*(4), 412–436. https://doi.org/10.1038/s41588-022-01024-z

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., & Pirinen, M. (2016). Finemap: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, *32*(10), 1493–1501. https://doi.org/10.1093/bioinformatics/btw018

Bennett, M. R., & Hacker, P. M. S. (2022, March 14). *Philosophical foundations of neuroscience*. John Wiley & Sons.

Bessadok, A., Mahjoub, M. A., & Rekik, I. (2022). Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(5), 5833–5848.

Bethlehem, R. a. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), 525–533. https://doi.org/10.1038/s41586-022-04554-y

Bi, Y., Abrol, A., Fu, Z., & Calhoun, V. D. (2023). A multimodal vision transformer for interpretable fusion of functional and structural neuroimaging data. https://doi.org/10.1101/2023.07.14.549002

Bijsterbosch, J., Harrison, S. J., Jbabdi, S., Woolrich, M., Beckmann, C., Smith, S., & Duff, E. P. (2020). Challenges and future directions for representations of functional brain organization. *Nature Neuroscience*, *23*(12), 1484–1495. https://doi.org/10.1038/s41593-020-00726-z

Bittner, N., Jockwitz, C., Franke, K., Gaser, C., Moebus, S., Bayen, U. J., Amunts, K., & Caspers, S. (2021). When your brain looks older than expected: Combined lifestyle risk and brainage. *Brain Structure and Function*, *226*(3), 621–645. https://doi.org/10.1007/s00429-020-02184-6

Bloch, L., & Friedrich, C. M. (2021). Comparison of automated volume extraction with freesurfer and fastsurfer for early alzheimer's disease detection with machine learning. *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, 113–118. https://doi.org/10.1109/CBMS52027.2021.00096

Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of

structural brain-behavior correlations. *Cortex*, *66*, 115–133. https://doi.org/10.1016/j.cortex.2014.11.019

Bogdan, R., Salmeron, B. J., Carey, C. E., Agrawal, A., Calhoun, V. D., Garavan, H., Hariri, A. R., Heinz, A., Hill, M. N., Holmes, A., Kalin, N. H., & Goldman, D. (2017). Imaging genetics and genomics in psychiatry: A critical review of progress and potential. *Biological Psychiatry*, *82*(3), 165–175. https://doi.org/10.1016/j.biopsych.2016.12.030

Böhle, M., Eitel, F., Weygandt, M., & Ritter, K. (2019). Layer-wise relevance propagation for explaining deep neural network decisions in mri-based alzheimer's disease classification. *Frontiers in Aging Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnagi.2019.00194

Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. *Journal of Experimental Psychology: General*, *151*(9), 2250–2258. https://doi.org/10.1037/xge0001181

Bookstein, F. L. (2001). "voxel-based morphometry" should not be used with imperfectly registered images. *NeuroImage*, *14*(6), 1454–1462. https://doi.org/10.1006/nimg.2001.0770

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Botvinik-Nezer, R., & Wager, T. D. (2022). Reproducibility in neuroimaging analysis: Challenges and solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2022.12.006

Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *International Journal of Epidemiology*, *44*(2), 512–525. https://doi.org/10.1093/ije/dyv080

Bowden, J., Davey Smith, G., Haycock, P. C., & Burgess, S. (2016). Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology*, *40*(4), 304–314. https://doi.org/10.1002/gepi.21965

Brainstorm Consortium, T., Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., … Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*(6395), eaap8757. https://doi.org/10.1126/science.aap8757

Brassington, J. C., & Marsh, N. V. (1998). Neuropsychological aspects of multiple sclerosis. *Neuropsychology Review*, *8*(2), 43–77. https://doi.org/10.1023/A:1025621700003

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Bron, E. E., Klein, S., Reinke, A., Papma, J. M., Maier-Hein, L., Alexander, D. C., & Oxtoby, N. P. (2022). Ten years of image analysis and machine learning competitions in dementia. *NeuroImage*, *253*, 119083.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Bulik-Sullivan, B. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Consortium, R., Consortium, P. G., for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, G. C., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., of the Psychiatric Genomics Consortium, S. W. G., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Busby, N., Newman-Norlund, S., Sayers, S., Newman-Norlund, R., Wilmskoetter, J., Rorden, C., Nemati, S., Wilson, S., Riccardi, N., Roth, R., Johnson, L., den Ouden, D. B., Fridriksson, J., & Bonilha, L. (2023).

Lower socioeconomic status is associated with premature brain aging. *Neurobiology of Aging*, *130*, 135–140. https://doi.org/10.1016/j.neurobiolaging.2023.06.012

Butler, E. R., Chen, A., Ramadan, R., Le, T. T., Ruparel, K., Moore, T. M., Satterthwaite, T. D., Zhang, F., Shou, H., Gur, R. C., Nichols, T. E., & Shinohara, R. T. (2021). Pitfalls in brain age analyses. *Human Brain Mapping*, *42*(13), 4092–4101. https://doi.org/10.1002/hbm.25533

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The uk biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00543

Bzdok, D., Engemann, D., & Thirion, B. (2020). Inference and prediction diverge in biomedicine. *Patterns*, *1*(8), 100119. https://doi.org/10.1016/j.patter.2020.100119

Bzdok, D., Varoquaux, G., & Steyerberg, E. W. (2021). Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry*, *78*(2), 127–128. https://doi.org/10.1001/jamapsychiatry.2020.2549

Bzdok, D., & Yeo, B. T. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, *155*, 549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061

Calegari, R., Ciatto, G., & Omicini, A. (2020). On the integration of symbolic and sub-symbolic techniques for xai: A survey. *Intelligenza Artificiale*, *14*(1), 7–32. https://doi.org/10.3233/IA-190036

Camacho, M., Wilms, M., Mouches, P., Almgren, H., Souza, R., Camicioli, R., Ismail, Z., Monchi, O., & Forkert, N. D. (2023). Explainable classification of parkinson's disease using deep learning trained on a large multicenter database of t1-weighted mri datasets. *NeuroImage: Clinical*, *38*, 103405. https://doi.org/10.1016/j.nicl.2023.103405

Cappelen, H., & Dever, J. (2021). *Making ai intelligible: Philosophical foundations*. Oxford University Press. https://library.oapen.org/handle/20.500.12657/48783

Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramèr, F., Balle, B., Ippolito, D., & Wallace, E. (2023). Extracting training data from diffusion models. *32nd USENIX Security Symposium (USENIX Security 23)*, 5253–5270. https://www.usenix.org/conference/usenixsecurity23/presentation/carlini

Carver, R. B., Fredheim, N. A. G., Mowinckel, A. M., Ebmeier, K. P., Friedman, B. B., Rosness, T. A., Drevon, C. A., Suri, S., Baaré, W. F. C., Zsoldos, E., Solé-Padullés, C., Bartrés-Faz, D., Ghisletta, P., Nawijn, L., Düzel, S., Madsen, K. S., Fjell, A. M., Lindenberger, U., Walhovd, K. B., & Budin-Ljøsne, I. (2022). People's interest in brain health testing: Findings from an international, online cross-sectional survey. *Frontiers in Public Health*, *10*. https://www.frontiersin.org/articles/10.3389/fpubh.2022.998302

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation plink: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Chang, D., Nalls, M. A., Hallgrímsdóttir, I. B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G. A., Ayalon, G., Bingol, B., Sheng, M., Hinds, D., Behrens, T. W., Singleton, A. B., Bhangale, T. R., & Graham, R. R. (2017). A meta-analysis of genome-wide association studies identifies 17 new parkinson's disease risk loci. *Nature Genetics*, *49*(10), 1511–1516. https://doi.org/10.1038/ng.3955

Chen, C.-H., Wang, Y., Lo, M.-T., Schork, A., Fan, C.-C., Holland, D., Kauppi, K., Smeland, O. B., Djurovic, S., Sanyal, N., Hibar, D. P., Thompson, P. M., Thompson, W. K., Andreassen, O. A., & Dale, A. M. (2017). Leveraging genome characteristics to improve gene discovery for putamen subcortical brain structure. *Scientific Reports*, *7*(1), 15736. https://doi.org/10.1038/s41598-017-15705-x

Chollet, F., et al. (2015). Keras. https://github.com/fchollet/keras

Cohen, J. W., Ramphal, B., DeSerisy, M., Zhao, Y., Pagliaccio, D., Colcombe, S., Milham, M. P., & Margolis, A. E. (2023). Relative brain age is associated with socioeconomic status and anxiety/depression problems in

youth. *Developmental Psychology*, No Pagination Specified–No Pagination Specified. https://doi.org/10.1037/dev0001593

Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S. E., Zhang, Q., Wray, N. R., Redmond, P., Marioni, R. E., Starr, J. M., Cox, S. R., Wardlaw, J. M., Sharp, D. J., & Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry*, *23*(5), 1385–1392. https://doi.org/10.1038/mp.2017.62

Cole, J. H. (2020). Multimodality neuroimaging brain-age in uk biobank: Relationship to biomedical, lifestyle, and cognitive factors. *Neurobiology of Aging*, *92*, 34–42. https://doi.org/10.1016/j.neurobiolaging.2020.03.014

Cole, J. H., & Franke, K. (2017). Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends in Neurosciences*, *40*(12), 681–690. https://doi.org/10.1016/j.tins.2017.10.001

Cole, J. H., Marioni, R. E., Harris, S. E., & Deary, I. J. (2019). Brain age and other bodily 'ages': Implications for neuropsychiatry. *Molecular Psychiatry*, *24*(2). https://doi.org/10.1038/s41380-018-0098-1

Cole, J. H., Raffel, J., Friede, T., Eshaghi, A., Brownlee, W. J., Chard, D., De Stefano, N., Enzinger, C., Pirpamer, L., Filippi, M., Gasperini, C., Rocca, M. A., Rovira, A., Ruggieri, S., Sastre-Garriga, J., Stromillo, M. L., Uitdehaag, B. M. J., Vrenken, H., Barkhof, F., … Group, s. M. (2020). Longitudinal assessment of multiple sclerosis with the brain-age paradigm. *Annals of Neurology*, *88*(1), 93–105. https://doi.org/10.1002/ana.25746

Dadi, K., Varoquaux, G., Houenou, J., Bzdok, D., Thirion, B., & Engemann, D. (2021). Population modeling with machine learning can enhance measures of mental health. *GigaScience*, *10*(10), giab071. https://doi.org/10.1093/gigascience/giab071

Davatzikos, C. (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, *23*(1), 17–20. https://doi.org/10.1016/j.neuroimage.2004.05.010

Davatzikos, C. (2019). Machine learning in neuroimaging: Progress and challenges. *NeuroImage*, *197*, 652–656. https://doi.org/10.1016/j.neuroimage.2018.10.003

Davey Smith, G., & Ebrahim, S. (2003). 'mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants

of disease?*. *International Journal of Epidemiology*, *32*(1), 1–22. https://doi.org/10.1093/ije/dyg070

Davidson-Pilon, C. (2019). Lifelines: Survival analysis in python. *Journal of Open Source Software*, *4*(40), 1317. https://doi.org/10.21105/joss.01317

de Filippis, R., Carbone, E. A., Gaetano, R., Bruni, A., Pugliese, V., Segura-Garcia, C., & De Fazio, P. (2019). Machine learning techniques in a structural and functional mri diagnostic approach in schizophrenia: A systematic review. *Neuropsychiatric Disease and Treatment*, *15*, 1605–1627. https://doi.org/10.2147/NDT.S202418

de Lange, A.-M. G., & Cole, J. H. (2020). Commentary: Correction procedures in brain-age prediction. *NeuroImage : Clinical*, *26*, 102229. https://doi.org/10.1016/j.nicl.2020.102229

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Dinga, R., Penninx, B. W., Veltman, D. J., Schmaal, L., & Marquand, A. F. (2019, August 22). *Beyond accuracy: Measures for assessing machine learning models, pitfalls and guidelines* (tech. rep.). Neuroscience. http://biorxiv.org/lookup/doi/10.1101/743138

Doan, N. T., Engvig, A., Zaske, K., Persson, K., Lund, M. J., Kaufmann, T., Cordova-Palomera, A., Alnæs, D., Moberget, T., Brækhus, A., Barca, M. L., Nordvik, J. E., Engedal, K., Agartz, I., Selbæk, G., Andreassen, O. A., Westlye, L. T., & Initiative, A. D. N. (2017). Distinguishing early and late brain aging from the alzheimer's disease spectrum: Consistent morphological patterns across independent samples. *NeuroImage*, *158*, 282–295. https://doi.org/10.1016/j.neuroimage.2017.06.070

Dörfel, R. P., Arenas-Gomez, J. M., Fisher, P. M., Ganz, M., Knudsen, G. M., Svensson, J., & Plavén-Sigray, P. (2023). Prediction of brain age using structural magnetic resonance imaging: A comparison of accuracy and test-retest reliability of publicly available software packages. https://doi.org/10.1101/2023.01.26.525514

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszko-

reit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. https://doi.org/10.48550/arXiv.2010.11929

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., & Scheltens, P. (2007). Research criteria for the diagnosis of alzheimer's disease: Revising the nincds–adrda criteria. *The Lancet Neurology*, *6*(8), 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3

Durães, F., Pinto, M., & Sousa, E. (2018). Old drugs as new treatments for neurodegenerative diseases. *Pharmaceuticals*, *11*(2), 44. https://doi.org/10.3390/ph11020044

Durstewitz, D., Koppe, G., & Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Molecular psychiatry*, *24*(11), 1583–1598.

Dyrba, M., Hanzig, M., Altenstein, S., Bader, S., Ballarini, T., Brosseron, F., Buerger, K., Cantré, D., Dechent, P., Dobisch, L., Düzel, E., Ewers, M., Fliessbach, K., Glanz, W., Haynes, J.-D., Heneka, M. T., Janowitz, D., Keles, D. B., Kilimann, I., … for the ADNI, A., DELCODE study groups. (2021). Improving 3d convolutional neural network comprehensibility via interactive visualization of relevance maps: Evaluation in alzheimer's disease. *Alzheimer's Research & Therapy*, *13*(1), 191. https://doi.org/10.1186/s13195-021-00924-2

Eickhoff, C. R., Hoffstaedter, F., Caspers, J., Reetz, K., Mathys, C., Dogan, I., Amunts, K., Schnitzler, A., & Eickhoff, S. B. (2021). Advanced brain ageing in parkinson's disease is related to disease duration and individual impairment. *Brain Communications*, *3*(3), fcab191. https://doi.org/10.1093/braincomms/fcab191

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *Neuroimage*, *59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718

Eickhoff, S. B., Nichols, T. E., Van Horn, J. D., & Turner, J. A. (2016). Sharing the wealth: Neuroimaging data repositories. *NeuroImage*, *124*(Pt B), 1065–1068. https://doi.org/10.1016/j.neuroimage.2015.10.079

Eitel, F., Soehler, E., Bellmann-Strobl, J., Brandt, A. U., Ruprecht, K., Giess, R. M., Kuchling, J., Asseyer, S., Weygandt, M., Haynes, J.-D., Scheel, M., Paul, F., & Ritter, K. (2019). Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional mri using layer-wise relevance propagation. *NeuroImage: Clinical*, *24*, 102003. https://doi.org/10.1016/j.nicl.2019.102003

Elliott, M. L., Belsky, D. W., Knodt, A. R., Ireland, D., Melzer, T. R., Poulton, R., Ramrakha, S., Caspi, A., Moffitt, T. E., & Hariri, A. R. (2019). Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. *Molecular Psychiatry*, 1–10. https://doi.org/10.1038/s41380-019-0626-7

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N. T., Lenzo, N., Martins, R. N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., … Group, A. R. (2009). The australian imaging, biomarkers and lifestyle (aibl) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. *International Psychogeriatrics*, *21*(4), 672–687. https://doi.org/10.1017/S1041610209009405

Emdin, C. A., Khera, A. V., & Kathiresan, S. (2017). Mendelian randomization. *JAMA*, *318*(19), 1925–1926. https://doi.org/10.1001/jama.2017.17219

Emre, M. (2003). Dementia associated with parkinson's disease. *The Lancet Neurology*, *2*(4), 229–237. https://doi.org/10.1016/S1474-4422(03)00351-X

Enache, D., Winblad, B., & Aarsland, D. (2011). Depression in dementia: Epidemiology, mechanisms, and treatment. *Current Opinion in Psychiatry*, *24*(6), 461. https://doi.org/10.1097/YCO.0b013e32834bb9d4

European Commision, Research, f. D.-G., Innovation, Baker, L., Cristea, I., Errington, T., Jaśko, K., Lusoli, W., MacCallum, C., Parry, V., Pérignon, C., Šimko, T., & Winchester, C. (2020). *Reproducibility of scientific results in the eu – scoping report* (W. Lusoli, Ed.). Publications Office.

Farahani, F. V., Fiok, K., Lahijanian, B., Karwowski, W., & Douglas, P. K. (2022). Explainable ai: A review of applications to neuroimaging data.

Frontiers in Neuroscience, *16*, 906290. https://doi.org/10.3389/fnins.2022.906290

Fathi, S., Ahmadi, M., & Dehnad, A. (2022). Early diagnosis of alzheimer's disease based on deep learning: A systematic review. *Computers in Biology and Medicine*, *146*, 105634. https://doi.org/10.1016/j.compbiomed.2022.105634

Ferrucci, L., Gonzalez-Freire, M., Fabbri, E., Simonsick, E., Tanaka, T., Moore, Z., Salimi, S., Sierra, F., & de Cabo, R. (2020). Measuring biological aging in humans: A quest. *Aging Cell*, *19*(2), e13080. https://doi.org/10.1111/acel.13080

Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., & Rocca, M. A. (2018). Multiple sclerosis. *Nature Reviews Disease Primers*, *4*(1), 1–27. https://doi.org/10.1038/s41572-018-0041-4

Fisch, L., Leenings, R., Winter, N. R., Dannlowski, U., Gaser, C., Cole, J. H., & Hahn, T. (2021). Editorial: Predicting chronological age from structural neuroimaging: The predictive analytics competition 2019. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyt.2021.710932

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11050–11055. https://doi.org/10.1073/pnas.200033797

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X

Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, *23 Suppl 1*, S69–84. https://doi.org/10.1016/j.neuroimage.2004.07.016

Fjell, A. M., & Walhovd, K. B. (2010). Structural brain changes in aging: Courses, causes and cognitive consequences. *Reviews in the Neurosciences*, *21*(3), 187–222. https://doi.org/10.1515/REVNEURO.2010.21.3.187

Floridi, L. (2013). *The philosophy of information*. OUP Oxford.

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*(3), 159–172. https://doi.org/10.1038/nrn3901

Fowler, C., Rainey-Smith, S. R., Bird, S., Bomke, J., Bourgeat, P., Brown, B. M., Burnham, S. C., Bush, A. I., Chadunow, C., Collins, S., Doecke, J., Doré, V., Ellis, K. A., Evered, L., Fazlollahi, A., Fripp, J., Gardener, S. L., Gibson, S., Grenfell, R., … Ames, D. (2021). Fifteen years of the australian imaging, biomarkers and lifestyle (aibl) study: Progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to alzheimer's disease. *Journal of Alzheimer's Disease Reports*, *5*(1), 443–468. https://doi.org/10.3233/ADR-210005

Franke, K., & Gaser, C. (2012). Longitudinal changes in individual brainage in healthy aging, mild cognitive impairment, and alzheimer's disease. *GeroPsych*, *25*(4), 235–245. https://doi.org/10.1024/1662-9647/a000074

Franke, K., & Gaser, C. (2019). Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained? *Frontiers in Neurology*, *10*. https://www.frontiersin.org/article/10.3389/fneur.2019.00789

Franke, K., Gaser, C., Manor, B., & Novak, V. (2013). Advanced brainage in older adults with type 2 diabetes mellitus. *Frontiers in Aging Neuroscience*, *5*. https://doi.org/10.3389/fnagi.2013.00090

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, *50*(3), 883–892. https://doi.org/10.1016/j.neuroimage.2010.01.005

Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., Tamminga, C. A., Gabbard, G. O., Gau, S. S.-F., Javitt, D. C., Oquendo, M. A., Shrout, P. E., Vieta, E., & Yager, J. (2013). The initial field trials of dsm-5: New blooms and old thorns. *American Journal of Psychiatry*, *170*(1), 1–5. https://doi.org/10.1176/appi.ajp.2012.12091189

Galvin, J. E., Pollack, J., & Morris, J. C. (2006). Clinical phenotype of parkinson disease dementia. *Neurology*, *67*(9), 1605–1611. https://doi.org/10.1212/01.wnl.0000242630.52203.8f

Gao, J., Chen, M., Xiao, D., Li, Y., Zhu, S., Li, Y., Dai, X., Lu, F., Wang, Z., Cai, S., & Wang, J. (2023). Classification of major depressive disorder using an attention-guided unified deep convolutional neural network

and individual structural covariance network. *Cerebral Cortex*, *33*(6), 2415–2425. https://doi.org/10.1093/cercor/bhac217

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*, *2*(11), 665–673. https://doi.org/10.1038/s42256-020-00257-z

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Goldstein, D. B. (2009). Common genetic variation and human traits. *New England Journal of Medicine*, *360*(17), 1696–1698. https://doi.org/10.1056/NEJMp0806284

Gong, W., Beckmann, C. F., Vedaldi, A., Smith, S. M., & Peng, H. (2021). Optimising a simple fully convolutional network for accurate brain age prediction in the pac 2019 challenge. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/article/10.3389/fpsyt.2021.627996

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., Ortega, B. N., Zaiko, Y. V., Roach, E. L., Korgaonkar, M. S., Grieve, S. M., Galatzer-Levy, I., Fox, P. T., & Etkin, A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA psychiatry*, *72*(4), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206

Guo, J., Yu, K., Dong, S.-S., Yao, S., Rong, Y., Wu, H., Zhang, K., Jiang, F., Chen, Y.-X., Guo, Y., & Yang, T.-L. (2022). Mendelian randomization analyses support causal relationships between brain imaging-derived phenotypes and risk of psychiatric disorders. *Nature Neuroscience*, *25*(11), 1519–1527. https://doi.org/10.1038/s41593-022-01174-7

Hahn, T., Ernsting, J., Winter, N. R., Holstein, V., Leenings, R., Beisemann, M., Fisch, L., Sarink, K., Emden, D., Opel, N., Redlich, R., Repple, J.,

Grotegerd, D., Meinert, S., Hirsch, J. G., Niendorf, T., Endemann, B., Bamberg, F., Kröncke, T., … Berger, K. (2022). An uncertainty-aware, shareable, and transparent neural network architecture for brain-age modeling. *Science Advances*, *8*(1), eabg9471. https://doi.org/10.1126/sciadv.abg9471

Hahn, T., Fisch, L., Ernsting, J., Winter, N. R., Leenings, R., Sarink, K., Emden, D., Kircher, T., Berger, K., & Dannlowski, U. (2021). From 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. *Brain*, *144*(3), e31–e31. https://doi.org/10.1093/brain/awaa454

Hajek, T., Franke, K., Kolenic, M., Capkova, J., Matejka, M., Propper, L., Uher, R., Stopkova, P., Novak, T., Paus, T., Kopecek, M., Spaniel, F., & Alda, M. (2019). Brain age in early stages of bipolar disorders or schizophrenia. *Schizophrenia Bulletin*, *45*(1), 190–198. https://doi.org/10.1093/schbul/sbx172

Han, L. K. M., Dinga, R., Hahn, T., Ching, C. R. K., Eyler, L. T., Aftanas, L., Aghajani, M., Aleman, A., Baune, B. T., Berger, K., Brak, I., Filho, G. B., Carballedo, A., Connolly, C. G., Couvy-Duchesne, B., Cullen, K. R., Dannlowski, U., Davey, C. G., Dima, D., … Schmaal, L. (2020). Brain aging in major depressive disorder: Results from the enigma major depressive disorder working group. *Molecular Psychiatry*, 1–16. https://doi.org/10.1038/s41380-020-0754-0

Hanson, J., Adkins, D., & Zhou, P. (2023, September 12). *Examining the reliability of brain age algorithms under varying degrees of subject motion* (tech. rep.). In Review. https://www.researchsquare.com/article/rs-3331689/v1

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with numpy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hauser, S. L., & Oksenberg, J. R. (2006). The neurobiology of multiple sclerosis: Genes, inflammation, and neurodegeneration. *Neuron*, *52*(1), 61–76. https://doi.org/10.1016/j.neuron.2006.09.011

Hayflick, L. (2004). Debates: The not-so-close relationship between biological aging and age-associated pathologies in humans. *The Journals of Geron-*

tology: Series A, *59*(6), B547–B550. https://doi.org/10.1093/gerona/59.6.B547

Hayflick, L. (2007). Biological aging is no longer an unsolved problem. *Annals of the New York Academy of Sciences*, *1100*, 1–13. https://doi.org/10.1196/annals.1395.001

He, T., Kong, R., Holmes, A. J., Nguyen, M., Sabuncu, M. R., Eickhoff, S. B., Bzdok, D., Feng, J., & Yeo, B. T. T. (2020). Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*, *206*, 116276. https://doi.org/10.1016/j.neuroimage.2019.116276

Heinrichs, J.-H. (2023). Brain age prediction and the challenge of biological concepts of aging. *Neuroethics*, *16*(3), 25. https://doi.org/10.1007/s12152-023-09531-4

Helzer, J. E., Kraemer, H. C., & Krueger, R. F. (2006). The feasibility and need for dimensional psychiatric diagnoses. *Psychological Medicine*, *36*(12), 1671–1680. https://doi.org/10.1017/S003329170600821X

Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., & Reuter, M. (2020). Fastsurfer - a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*, *219*, 117012. https://doi.org/10.1016/j.neuroimage.2020.117012

Herzog, C. (2022). On the ethical and epistemological utility of explicable ai in medicine. *Philosophy & Technology*, *35*(2), 50. https://doi.org/10.1007/s13347-022-00546-y

Hibar, D. P., Westlye, L. T., Doan, N. T., Jahanshad, N., Cheung, J. W., Ching, C. R. K., Versace, A., Bilderbeck, A. C., Uhlmann, A., Mwangi, B., Krämer, B., Overs, B., Hartberg, C. B., Abé, C., Dima, D., Grotegerd, D., Sprooten, E., Bøen, E., Jimenez, E., … Andreassen, O. A. (2018). Cortical abnormalities in bipolar disorder: An mri analysis of 6503 individuals from the enigma bipolar disorder working group. *Molecular Psychiatry*, *23*(4), 932–942. https://doi.org/10.1038/mp.2017.73

Hibar, D. P., Westlye, L. T., van Erp, T. G. M., Rasmussen, J., Leonardo, C. D., Faskowitz, J., Haukvik, U. K., Hartberg, C. B., Doan, N. T., Agartz, I., Dale, A. M., Gruber, O., Krämer, B., Trost, S., Liberg, B., Abé, C., Ekman, C. J., Ingvar, M., Landén, M., … Andreassen, O. A. (2016). Subcortical volumetric abnormalities in bipolar disorder. *Molecular Psychiatry*, *21*(12), 1710–1716. https://doi.org/10.1038/mp.2015.227

Hinshaw, S. P., & Stier, A. (2008). Stigma as related to mental disorders. *Annual Review of Clinical Psychology*, *4*(1), 367–393. https://doi.org/10.1146/annurev.clinpsy.4.022007.141245

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–434. https://doi.org/10.1016/j.tics.2007.09.004

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Høgestøl, E. A., Kaufmann, T., Nygaard, G. O., Beyer, M. K., Sowa, P., Nordvik, J. E., Kolskår, K., Richard, G., Andreassen, O. A., Harbo, H. F., & Westlye, L. T. (2019). Cross-sectional and longitudinal mri brain scans reveal accelerated brain aging in multiple sclerosis. *Frontiers in Neurology*, *0*. https://doi.org/10.3389/fneur.2019.00450

Holdcroft, A. (2007). Gender bias in research: How does it affect evidence based medicine? *Journal of the Royal Society of Medicine*, *100*(1), 2–3.

Holliday, R. (2004). Debates: The close relationship between biological aging and age-associated pathologies in humans. *The Journals of Gerontology: Series A*, *59*(6), B543–B546. https://doi.org/10.1093/gerona/59.6.B543

Holm, M. C., Leonardsen, E. H., Beck, D., Dahl, A., Kjelkenes, R., Lange, d. A.-M. G., & Westlye, L. T. (2023). Linking brain maturation and puberty during early adolescence using longitudinal brain age prediction in the abcd cohort. *Developmental Cognitive Neuroscience*, *60*, 101220. https://doi.org/https://doi.org/10.1016/j.dcn.2023.101220

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, *9*(4), e1312. https://doi.org/10.1002/widm.1312

Hong, J., Huang, Y., Ye, J., Wang, J., Xu, X., Wu, Y., Li, Y., Zhao, J., Li, R., Kang, J., & Lai, X. (2022). 3d frn-resnet: An automated major depressive disorder structural magnetic resonance imaging data identification framework. *Frontiers in Aging Neuroscience*, *14*. https://www.frontiersin.org/articles/10.3389/fnagi.2022.912283

Horien, C., Noble, S., Greene, A. S., Lee, K., Barron, D. S., Gao, S., O'Connor, D., Salehi, M., Dadashkarimi, J., Shen, X., Lake, E. M. R., Constable, R. T., & Scheinost, D. (2021). A hitchhiker's guide to working with

large, open-source neuroimaging datasets. *Nature Human Behaviour*, *5*(2), 185–193. https://doi.org/10.1038/s41562-020-01005-4

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Iglesias, J. E., Billot, B., Balbastre, Y., Magdamo, C., Arnold, S. E., Das, S., Edlow, B. L., Alexander, D. C., Golland, P., & Fischl, B. (2023). Synthsr: A public ai tool to turn heterogeneous clinical brain scans into high-resolution t1-weighted images for 3d morphometry. *Science Advances*, *9*(5), eadd3607. https://doi.org/10.1126/sciadv.add3607

Ioannidis, J. P. A. (2011). Excess significance bias in the literature on brain volume abnormalities. *Archives of General Psychiatry*, *68*(8), 773–780. https://doi.org/10.1001/archgenpsychiatry.2011.28

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., … Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. *Nature Genetics*, *51*(3), 404–413. https://doi.org/10.1038/s41588-018-0311-9

Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 798–808. https://doi.org/10.1016/j.bpsc.2018.04.004

Japkowicz, N. (2013). Assessment metrics for imbalanced learning. In *Imbalanced learning* (pp. 187–206). John Wiley & Sons, Ltd. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118646106.ch8

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8

Jin, T., Nguyen, N. D., Talos, F., & Wang, D. (2021). ECMarker: Interpretable machine learning model identifies gene expression biomarkers predicting clinical outcomes and reveals molecular mechanisms of human disease

in early stages. *Bioinformatics (Oxford, England)*, *37*(8), 1115–1124. https://doi.org/10.1093/bioinformatics/btaa935

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, *15*(4), 673–692. https://doi.org/10.1016/j.concog.2006.09.004

Jonsson, B. A., Bjornsdottir, G., Thorgeirsson, T. E., Ellingsen, L. M., Walters, G. B., Gudbjartsson, D. F., Stefansson, H., Stefansson, K., & Ulfarsson, M. O. (2019). Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications*, *10*(1), 5409. https://doi.org/10.1038/s41467-019-13163-9

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kahn, R. S., Sommer, I. E., Murray, R. M., Meyer-Lindenberg, A., Weinberger, D. R., Cannon, T. D., O'Donovan, M., Correll, C. U., Kane, J. M., van Os, J., & Insel, T. R. (2015). Schizophrenia. *Nature Reviews Disease Primers*, *1*(1), 1–23. https://doi.org/10.1038/nrdp.2015.67

Karani, N., Chaitanya, K., Baumgartner, C., & Konukoglu, E. (2018). A lifelong learning approach to brain mr segmentation across scanners and protocols. In A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-030-00928-1_54

Kaufmann, T., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I., Alnæs, D., Barch, D. M., Baur-Streubel, R., Bertolino, A., Bettella, F., Beyer, M. K., Bøen, E., Borgwardt, S., Brandt, C. L., Buitelaar, J., Celius, E. G., Cervenka, S., Conzelmann, A., … Westlye, L. T. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, *22*(10), 1617–1623. https://doi.org/10.1038/s41593-019-0471-7

Keefe, R. S. E., & Harvey, P. D. (2012). Cognitive impairment in schizophrenia. In M. A. Geyer & G. Gross (Eds.), *Novel antischizophrenia treatments* (pp. 11–37). Springer. https://doi.org/10.1007/978-3-642-25758-2_2

Khorrami, P., Paine, T. L., & Huang, T. S. (2015). Do deep neural networks learn facial action units when doing expression recognition? *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, 19–27. https://doi.org/10.1109/ICCVW.2015.12

Kindermans, P.-J., Hooker, S., Adebayo, J., Alber, M., Schütt, K. T., Dähne, S., Erhan, D., & Kim, B. (2019). The (un)reliability of saliency methods. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable ai: Interpreting, explaining and visualizing deep learning* (pp. 267–280). Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_14

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. https://doi.org/10.48550/arXiv.1609.02907

Knopman, D. S., Amieva, H., Petersen, R. C., Chételat, G., Holtzman, D. M., Hyman, B. T., Nixon, R. A., & Jones, D. T. (2021). Alzheimer disease. *Nature Reviews Disease Primers*, *7*(1), 1–21. https://doi.org/10.1038/s41572-021-00269-y

Kohlbrenner, M., Bauer, A., Nakajima, S., Binder, A., Samek, W., & Lapuschkin, S. (2020). Towards best practice in explaining neural network decisions with lrp. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–7. https://doi.org/10.1109/IJCNN48605.2020.9206975

Kopal, J., & Bzdok, D. (2023). Endorsing complexity through diversity: Computational psychiatry meets big data analytics. *Biological Psychiatry*, *93*(8), 655–657. https://doi.org/10.1016/j.biopsych.2022.07.023

Korbmacher, M., Gurholt, T. P., de Lange, A.-M. G., van der Meer, D., Beck, D., Eikefjord, E., Lundervold, A., Andreassen, O. A., Westlye, L. T., & Maximov, I. I. (2023). Bio-psycho-social factors' associations with brain age: A large-scale uk biobank diffusion study of 35,749 participants. *Frontiers in Psychology*, *14*. https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1117732

Korbmacher, M., Wang, M.-Y., Eikeland, R., Buchert, R., Andreassen, O. A., Espeseth, T., Leonardsen, E., Westlye, L. T., Maximov, I. I., & Specht, K. (2023). Considerations on brain age predictions from repeatedly sampled data across time. *Brain and Behavior*, *n/a*(n/a), e3219. https://doi.org/https://doi.org/10.1002/brb3.3219

Korda, A. I., Ruef, A., Neufang, S., Davatzikos, C., Borgwardt, S., Meisenzahl, E. M., & Koutsouleris, N. (2021). Identification of voxel-based texture

abnormalities as new biomarkers for schizophrenia and major depressive patients using layer-wise relevance propagation on deep learning decisions. *Psychiatry Research. Neuroimaging*, *313*, 111303. https://doi.org/10.1016/j.pscychresns.2021.111303

Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T., Falkai, P., Riecher-Rössler, A., Möller, H.-J., Reiser, M., Pantelis, C., & Meisenzahl, E. (2014). Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders. *Schizophrenia Bulletin*, *40*(5), 1140–1153. https://doi.org/10.1093/schbul/sbt142

Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry*, *157*(2), 163–171. https://doi.org/10.1176/appi.ajp.157.2.163

Kundu, S. (2021). Ai in medicine must be explainable. *Nature Medicine*, *27*(8), 1328–1328. https://doi.org/10.1038/s41591-021-01461-z

Kupfer, D. J., & Regier, D. A. (2010). Why all of medicine should care about dsm-5. *JAMA*, *303*(19), 1974–1975. https://doi.org/10.1001/jama.2010.646

Kwak, S., Kim, H., Chey, J., & Youm, Y. (2018). Feeling how old i am: Subjective age is associated with estimated brain age. *Frontiers in Aging Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnagi.2018.00168

Laansma, M. A., Bright, J. K., Al-Bachari, S., Anderson, T. J., Ard, T., Assogna, F., Baquero, K. A., Berendse, H. W., Blair, J., Cendes, F., Dalrymple-Alford, J. C., de Bie, R. M., Debove, I., Dirkx, M. F., Druzgal, J., Emsley, H. C., Garraux, G., Guimarães, R. P., Gutman, B. A., … Study, T. E.-P. (2021). International multicenter analysis of brain structure across clinical stages of parkinson's disease. *Movement Disorders*, *36*(11), 2583–2594. https://doi.org/10.1002/mds.28706

Laird, A. R., Lancaster, J. L., & Fox, P. T. (2005). Brainmap: The social evolution of a human brain mapping database. *Neuroinformatics*, *3*(1), 65–78. https://doi.org/10.1385/ni:3:1:065

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., Raichle, M. E., Cruchaga, C., & Marcus, D. (2019). Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. https://doi.org/10.1101/2019.12.13.19014902

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. In D. Touretzky (Ed.). Morgan-Kaufmann. https://proceedings.neurips.cc/paper_files/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf

Lee, C. H., Cook, S., Lee, J. S., & Han, B. (2016). Comparison of two meta-analysis methods: Inverse-variance-weighted average and weighted sum of z-scores. *Genomics & Informatics*, *14*(4), 173–180. https://doi.org/10.5808/GI.2016.14.4.173

Lee, P. H., Feng, Y.-C. A., & Smoller, J. W. (2021). Pleiotropy and cross-disorder genetics among psychiatric disorders. *Biological Psychiatry*, *89*(1), 20–31. https://doi.org/10.1016/j.biopsych.2020.09.026

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, *21*(1), 133–145. https://doi.org/10.1002/wps.20941

Liang, Z.-P., & Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging*. SPIE Optical Engineering Press Belllingham, WA.

Lin, Q.-H., Niu, Y.-W., Sui, J., Zhao, W.-D., Zhuo, C., & Calhoun, V. D. (2022). Sspnet: An interpretable 3d-cnn for classification of schizophrenia using phase maps of resting-state complex-valued fmri data. *Medical Image Analysis*, *79*, 102430. https://doi.org/10.1016/j.media.2022.102430

Linli, Z., Feng, J., Zhao, W., & Guo, S. (2022). Associations between smoking and accelerated brain ageing. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *113*, 110471. https://doi.org/10.1016/j.pnpbp.2021.110471

Lipton, Z. C. (2017, March 6). The mythos of model interpretability. http://arxiv.org/abs/1606.03490

Liu, Y., Sangineto, E., Bi, W., Sebe, N., Lepri, B., & Nadai, M. (2021). Efficient training of visual transformers with small datasets. *Advances in Neural Information Processing Systems*, *34*, 23818–23830. https://proceedings.neurips.cc/paper/2021/hash/c81e155d85dae5430a8cee6f2242e82c-Abstract.html

Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L.-O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., & Consortium, A. (2011). Combination analysis of neuropsychological tests and structural mri measures in differentiating ad, mci and control groups–the addneuromed study. *Neurobiology of Aging*, *32*(7), 1198–1206. https://doi.org/10.1016/j.neurobiolaging.2009.07.008

Lo, M.-T., Wang, Y., Kauppi, K., Sanyal, N., Fan, C.-C., Smeland, O. B., Schork, A., Holland, D., Hinds, D. A., Tung, J. Y., Andreassen, O. A., Dale, A. M., & Chen, C.-H. (2017). Modeling prior information of common genetic variants improves gene discovery for neuroticism. *Human Molecular Genetics*, *26*(22), 4530–4539. https://doi.org/10.1093/hmg/ddx340

London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Center Report*, *49*(1), 15–21. https://doi.org/10.1002/hast.973

López-Rubio, E. (2021). Throwing light on black boxes: Emergence of visual categories from deep learning. *Synthese*, *198*(10), 10021–10041. https://doi.org/10.1007/s11229-020-02700-5

Loth, E., Ahmad, J., Chatham, C., López, B., Carter, B., Crawley, D., Oakley, B., Hayward, H., Cooke, J., Cáceres, A. S. J., Bzdok, D., Jones, E., Charman, T., Beckmann, C., Bourgeron, T., Toro, R., Buitelaar, J., Murphy, D., & Dumas, G. (2021). The meaning of significant mean group differences for biomarker discovery. *PLOS Computational Biology*, *17*(11), e1009477. https://doi.org/10.1371/journal.pcbi.1009477

Lovestone, S., Francis, P., & Strandgaard, K. (2007). Biomarkers for disease modification trials–the innovative medicines initiative and addneuromed. *The Journal of Nutrition, Health & Aging*, *11*(4), 359–361.

Löwe, L. C., Gaser, C., Franke, K., & Initiative, A. D. N. (2016). The effect of the apoe genotype on individual brainage in normal aging, mild cognitive impairment, and alzheimer's disease. *PloS One*, *11*(7), e0157514. https://doi.org/10.1371/journal.pone.0157514

Madan, C. R. (2022). Scan once, analyse many: Using large open-access neuroimaging datasets to understand the brain. *Neuroinformatics*, *20*(1), 109–137. https://doi.org/10.1007/s12021-021-09519-6

Magesh, P. R., Myloth, R. D., & Tom, R. J. (2020). An explainable machine learning model for early detection of parkinson's disease using lime

on datscan imagery. *Computers in Biology and Medicine*, *126*, 104041. https://doi.org/10.1016/j.compbiomed.2020.104041

Malone, I. B., Cash, D., Ridgway, G. R., MacManus, D. G., Ourselin, S., Fox, N. C., & Schott, J. M. (2013). Miriad–public release of a multiple time point alzheimer's mr imaging dataset. *NeuroImage*, *70*, 33–36. https://doi.org/10.1016/j.neuroimage.2012.12.044

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9

Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Golland, P., Klein, S., & Alexander, D. C. (2019). Tadpole challenge: Accurate alzheimer's disease prediction through crowdsourced forecasting of future data. In I. Rekik, E. Adeli, & S. H. Park (Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-030-32281-6_1

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, *137*(5), 856–879. https://doi.org/10.1037/a0023678

Marquand, A. F., Kia, S. M., Zabihi, M., Wolfers, T., Buitelaar, J. K., & Beckmann, C. F. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry*, *24*(10), 1415–1424. https://doi.org/10.1038/s41380-019-0441-1

Marquand, A. F., Rezek, I., Buitelaar, J., & Beckmann, C. F. (2016). Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies. *Biological psychiatry*, *80*(7), 552–561.

Martin, S. A., Townend, F. J., Barkhof, F., & Cole, J. H. (2023). Interpretable machine learning for dementia: A systematic review. *Alzheimer's & Dementia*, *19*(5), 2135–2149. https://doi.org/10.1002/alz.12948

Martyn, P., McPhilemy, G., Nabulsi, L., Martyn, F., McDonald, C., Cannon, D. M., & Schukat, M. (2019). Using magnetic resonance imaging to distinguish a healthy brain from a bipolar brain: A transfer learning approach. *AICS*, 16–27.

Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., & Evans, A. C. (2018). Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage: Clinical*, *20*, 506–522. https://doi.org/10.1016/j.nicl.2018.08.019

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of parkinson's disease: A review of literature. *Frontiers in Aging Neuroscience*, *13*. https://www.frontiersin.org/articles/10.3389/fnagi.2021.633752

Milham, M., Fair, D., Mennes, M., & Mostofsky, S. (2012). The adhd-200 consortium: A model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in Systems Neuroscience*, *6*, 62. https://doi.org/10.3389/fnsys.2012.00062

Millard, L. A., Davies, N. M., Gaunt, T. R., Davey Smith, G., & Tilling, K. (2018). Software application profile: Phesant: A tool for performing automated phenome scans in uk biobank. *International Journal of Epidemiology*, *47*(1), 29–35. https://doi.org/10.1093/ije/dyx204

Mirzaei, G., & Adeli, H. (2022). Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia. *Biomedical Signal Processing and Control*, *72*, 103293. https://doi.org/10.1016/j.bspc.2021.103293

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford university press.

Moazami, F., Lefevre-Utile, A., Papaloukas, C., & Soumelis, V. (2021). Machine learning approaches in study of multiple sclerosis disease through magnetic resonance images. *Frontiers in Immunology*, *12*. https://www.frontiersin.org/articles/10.3389/fimmu.2021.700582

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K.-R. (2019). Layer-wise relevance propagation: An overview. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable ai: Interpreting, explaining and visualizing deep learning* (pp. 193–209). Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_10

Moridi, T., Stawiarz, L., McKay, K. A., Ineichen, B. V., Ouellette, R., Ferreira, D., Muehlboeck, J.-S., Westman, E., Kockum, I., Olsson, T., Piehl, F.,

Hillert, J., Manouchehrinia, A., & Granberg, T. (2022). Association between brain volume and disability over time in multiple sclerosis. *Multiple Sclerosis Journal - Experimental, Translational and Clinical*, *8*(4), 20552173221144230. https://doi.org/10.1177/20552173221144230

Morris, J. C., Weintraub, S., Chui, H. C., Cummings, J., Decarli, C., Ferris, S., Foster, N. L., Galasko, D., Graff-Radford, N., Peskind, E. R., Beekly, D., Ramos, E. M., & Kukull, W. A. (2006). The uniform data set (uds): Clinical and cognitive variables and descriptive data from alzheimer disease centers. *Alzheimer Disease and Associated Disorders*, *20*(4), 210–216. https://doi.org/10.1097/01.wad.0000213865.09806.92

Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M., & He, X. (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics*, *52*(7), 740–747. https://doi.org/10.1038/s41588-020-0631-4

Moynihan, R., Doust, J., & Henry, D. (2012). Preventing overdiagnosis: How to stop harming the healthy. *BMJ*, *344*(may28 4), e3502–e3502. https://doi.org/10.1136/bmj.e3502

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R. I., Qiao, Z., Als, T. D., Bigdeli, T. B., Børte, S., Bryois, J., Charney, A. W., Drange, O. K., Gandal, M. J., Hagenaars, S. P., Ikeda, M., Kamitaki, N., Kim, M., Krebs, K., Panagiotaropoulou, G., … Andreassen, O. A. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*, *53*(6), 817–829. https://doi.org/10.1038/s41588-021-00857-4

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, *12*(2), 229–244. https://doi.org/10.1007/s12021-013-9204-3

Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D. A., Noyce, A. J., Xue, A., Bras, J., Young, E., von Coelln, R., Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrom, L., Siitonen, A., … Singleton, A. B. (2019). Identification of novel risk loci, causal insights, and heritable risk for parkinson's disease: A meta-genome wide association study. *The Lancet. Neurology*, *18*(12), 1091–1102. https://doi.org/10.1016/S1474-4422(19)30320-5

Nenadić, I., Dietzek, M., Langbein, K., Sauer, H., & Gaser, C. (2017). Brainage score indicates accelerated brain aging in schizophrenia, but not bipolar

disorder. *Psychiatry Research: Neuroimaging*, *266*, 86–89. https://doi.org/10.1016/j.pscychresns.2017.05.006

Nesvåg, R., Jönsson, E. G., Bakken, I. J., Knudsen, G. P., Bjella, T. D., Reichborn-Kjennerud, T., Melle, I., & Andreassen, O. A. (2017). The quality of severe mental disorder diagnoses in a national health registry as compared to research diagnoses based on structured interview. *BMC Psychiatry*, *17*, 93. https://doi.org/10.1186/s12888-017-1256-8

Ning, K., Zhao, L., Matloff, W., Sun, F., & Toga, A. W. (2020). Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants. *Scientific Reports*, *10*(1), 10. https://doi.org/10.1038/s41598-019-56089-4

Nishino, J., Ochi, H., Kochi, Y., Tsunoda, T., & Matsui, S. (2018). Sample size for successful genome-wide association study of major depressive disorder. *Frontiers in Genetics*, *9*. https://www.frontiersin.org/articles/10.3389/fgene.2018.00227

Noor, M. B. T., Zenia, N. Z., Kaiser, M. S., Mamun, S. A., & Mahmud, M. (2020). Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of alzheimer's disease, parkinson's disease and schizophrenia. *Brain Informatics*, *7*(1), 11. https://doi.org/10.1186/s40708-020-00112-2

OpenAI. (2015, December 11). Introducing openai. https://openai.com/blog/introducing-openai

OpenAI. (2019, March 11). Openai lp. https://openai.com/blog/openai-lp

OpenAI. (2020, July 6). Openai api. https://web.archive.org/web/20200706151747/https://openai.com/blog/openai-api/

OpenAI. (2022, November 30). Introducing chatgpt. https://openai.com/blog/chatgpt

OpenAI. (2023, February 1). Introducing chatgpt plus. https://openai.com/blog/chatgpt-plus

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., Mohr, D. C., & Schatzberg, A. F. (2016). Major depressive disorder. *Nature Reviews Disease Primers*, *2*(1), 1–20. https://doi.org/10.1038/nrdp.2016.65

Páez, A. (2019). The pragmatic turn in explainable artificial intelligence (xai). *Minds and Machines*, *29*(3), 441–459. https://doi.org/10.1007/s11023-019-09502-w

Pan, Z., Park, C., Brietzke, E., Zuckerman, H., Rong, C., Mansur, R. B., Fus, D., Subramaniapillai, M., Lee, Y., & McIntyre, R. S. (2019). Cognitive impairment in major depressive disorder. *CNS Spectrums*, *24*(1), 22–29. https://doi.org/10.1017/S1092852918001207

Pearlson, G. D. (2015). Etiologic, phenomenologic, and endophenotypic overlap of schizophrenia and bipolar disorder. *Annual Review of Clinical Psychology*, *11*(1), 251–281. https://doi.org/10.1146/annurev-clinpsy-032814-112915

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., & Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, *68*, 101871. https://doi.org/10.1016/j.media.2020.101871

Persson, K., Leonardsen, E. H., Edwin, T. H., Knapskog, A.-B., Tangen, G. G., Selbæk, G., Wolfers, T., Westlye, L. T., & Engedal, K. (2023). Diagnostic accuracy of brain age prediction in a memory clinic population and comparison with clinically available volumetric measures. *Scientific Reports*, *13*(1), 14957. https://doi.org/10.1038/s41598-023-42354-0

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Weiner, M. W. (2010). Alzheimer's disease neuroimaging initiative (adni). *Neurology*, *74*(3), 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25

Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., Johnson, H. J., Paulsen, J. S., Turner, J. A., & Calhoun, V. D. (2014). Deep learning for neuroimaging: A validation study. *Frontiers in Neuroscience*, *8*, 229. https://doi.org/10.3389/fnins.2014.00229

Plis, S. M., Sarwate, A. D., Wood, D., Dieringer, C., Landis, D., Reed, C., Panta, S. R., Turner, J. A., Shoemaker, J. M., Carter, K. W., Thompson, P., Hutchison, K., & Calhoun, V. D. (2016). Coinstac: A privacy enabled

model and prototype for leveraging and processing decentralized brain imaging data. *Frontiers in Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnins.2016.00365

Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., Schrag, A.-E., & Lang, A. E. (2017). Parkinson disease. *Nature Reviews Disease Primers*, *3*(1), 1–21. https://doi.org/10.1038/nrdp.2017.13

Quaak, M., van de Mortel, L., Thomas, R. M., & van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis. *NeuroImage: Clinical*, *30*, 102584. https://doi.org/10.1016/j.nicl.2021.102584

Rahman, M. M., Calhoun, V. D., & Plis, S. M. (2023). Looking deeper into interpretable deep learning in neuroimaging: A comprehensive survey. https://doi.org/10.48550/arXiv.2307.09615

Rahman, M. M., Mahmood, U., Lewis, N., Gazula, H., Fedorov, A., Fu, Z., Calhoun, V. D., & Plis, S. M. (2022). Interpreting models interpreting brain dynamics. *Scientific Reports*, *12*(1), 12023. https://doi.org/10.1038/s41598-022-15539-2

Rando, T. A., & Chang, H. Y. (2012). Aging, rejuvenation, and epigenetic reprogramming: Resetting the aging clock. *Cell*, *148*(1), 46–57. https://doi.org/10.1016/j.cell.2012.01.003

Rasmussen, J., & Langerman, H. (2019). Alzheimer's disease – why we need early diagnosis. *Degenerative Neurological and Neuromuscular Disease*, *9*, 123–130. https://doi.org/10.2147/DNND.S228939

Relton, C. L., & Davey Smith, G. (2012). Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *International Journal of Epidemiology*, *41*(1), 161–176. https://doi.org/10.1093/ije/dyr233

Ricard, J. A., Parker, T. C., Dhamala, E., Kwasa, J., Allsop, A., & Holmes, A. J. (2023). Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data. *Nature Neuroscience*, *26*(1), 4–11. https://doi.org/10.1038/s41593-022-01218-y

Risacher, S. L., & Saykin, A. J. (2013). Neuroimaging biomarkers of neurodegenerative diseases and dementia. *Seminars in Neurology*, *33*(4), 386–416. https://doi.org/10.1055/s-0033-1359312

Robinson, E. C., Hammers, A., Ericsson, A., Edwards, A. D., & Rueckert, D. (2010). Identifying population differences in whole-brain structural networks: A machine learning approach. *NeuroImage*, *50*(3), 910–919.

Robinson, L. J., & Nicol Ferrier, I. (2006). Evolution of cognitive impairment in bipolar disorder: A systematic review of cross-sectional evidence. *Bipolar Disorders*, *8*(2), 103–116. https://doi.org/10.1111/j.1399-5618.2006.00277.x

Roelfs, D., Frei, O., van der Meer, D., Tissink, E., Shadrin, A., Alnaes, D., Andreassen, O. A., Westlye, L. T., & Kaufmann, T. (2023). Shared genetic architecture between mental health and the brain functional connectome in the uk biobank. *BMC Psychiatry*, *23*(1), 461. https://doi.org/10.1186/s12888-023-04905-7

Rokicki, J., Wolfers, T., Nordhøy, W., Tesli, N., Quintana, D. S., Alnæs, D., Richard, G., Lange, d. A.-M. G., Lund, M. J., Norbom, L., Agartz, I., Melle, I., Nærland, T., Selbæk, G., Persson, K., Nordvik, J. E., Schwarz, E., Andreassen, O. A., Kaufmann, T., & Westlye, L. T. (2021). Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Human Brain Mapping*, *42*(6), 1714–1726. https://doi.org/10.1002/hbm.25323

Rolls, E. T. (2021). A neuroscience levels of explanation approach to the mind and the brain. *Frontiers in Computational Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fncom.2021.649679

Rootes-Murdy, K., Gazula, H., Verner, E., Kelly, R., DeRamus, T., Plis, S., Sarwate, A., Turner, J., & Calhoun, V. (2022). Federated analysis of neuroimaging data: A review of the field. *Neuroinformatics*, *20*(2), 377–390. https://doi.org/10.1007/s12021-021-09550-7

Rosenberg, M. D., Casey, B. J., & Holmes, A. J. (2018). Prediction complements explanation in understanding the developing brain. *Nature Communications*, *9*(1), 589. https://doi.org/10.1038/s41467-018-02887-9

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Rutherford, S., Fraza, C., Dinga, R., Kia, S. M., Wolfers, T., Zabihi, M., Berthet, P., Worker, A., Verdi, S., Andrews, D., Han, L. K., Bayer, J. M., Dazzan, P., McGuire, P., Mocking, R. T., Schene, A., Sripada, C., Tso, I. F.,

Duval, E. R., … Marquand, A. F. (2022). Charting brain growth and aging at high spatial precision (C. I. Baker, B. Taschler, O. Esteban, & T. Constable, Eds.). *eLife*, *11*, e72904. https://doi.org/10.7554/eLife.72904

Saghayan, M. H., Zolfagharnasab, M. H., Khadem, A., Matinfar, F., & Rashidi, H. (2023). Diagnosing bipolar disorder from 3-d structural magnetic resonance images using a hybrid gan-cnn method. https://doi.org/10.48550/arXiv.2310.07359

Saglam, Y., Oz, A., Yildiz, G., Ermis, C., Kargin, O. A., Arslan, S., & Karacetin, G. (2023). Can diffusion tensor imaging have a diagnostic utility to differentiate early-onset forms of bipolar disorder and schizophrenia: A neuroimaging study with explainable machine learning algorithms. *Psychiatry Research: Neuroimaging*, *335*, 111696. https://doi.org/10.1016/j.pscychresns.2023.111696

Samek, W., Binder, A., Montavon, G., Lapuschkin, S., & Müller, K.-R. (2017). Evaluating the visualization of what a deep neural network has learned. *IEEE Transactions on Neural Networks and Learning Systems*, *28*, 2660–2673. https://doi.org/10.1109/TNNLS.2016.2599820

Samek, W., & Müller, K.-R. (2019). Towards explainable artificial intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable ai: Interpreting, explaining and visualizing deep learning* (pp. 5–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_1

Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., Palmer, T., Schooling, C. M., Wallace, C., Zhao, Q., & Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, *2*(1), 1–21. https://doi.org/10.1038/s43586-021-00092-5

Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, *19*(8), 491–504. https://doi.org/10.1038/s41576-018-0016-z

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M., Salehi, M., Gao, S., Shen, X., O'Connor, D., Barron, D. S., Yip, S. W., Rosenberg, M. D., & Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, *193*, 35–45. https://doi.org/10.1016/j.neuroimage.2019.02.057

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., Cheung, J. W., van Erp, T. G. M., Bos, D., Ikram, M. A., Vernooij,

M. W., Niessen, W. J., Tiemeier, H., Hofman, A., Wittfeld, K., Grabe, H. J., Janowitz, D., Bülow, R., Selonke, M., … Veltman, D. J. (2017). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the enigma major depressive disorder working group. *Molecular Psychiatry*, *22*(6), 900–909. https://doi.org/10.1038/mp.2016.60

Schmaal, L., Pozzi, E., C. Ho, T., van Velzen, L. S., Veer, I. M., Opel, N., Van Someren, E. J. W., Han, L. K. M., Aftanas, L., Aleman, A., Baune, B. T., Berger, K., Blanken, T. F., Capitão, L., Couvy-Duchesne, B., R. Cullen, K., Dannlowski, U., Davey, C., Erwin-Grabner, T., … Veltman, D. J. (2020). Enigma mdd: Seven years of global neuroimaging studies of major depression through worldwide data sharing. *Translational Psychiatry*, *10*(1), 1–19. https://doi.org/10.1038/s41398-020-0842-6

Schnack, H. G., van Haren, N. E., Nieuwenhuis, M., Hulshoff Pol, H. E., Cahn, W., & Kahn, R. S. (2016). Accelerated brain aging in schizophrenia: A longitudinal pattern recognition study. *American Journal of Psychiatry*, *173*(6), 607–616. https://doi.org/10.1176/appi.ajp.2015.15070922

Schulz, M.-A., Yeo, B. T. T., Vogelstein, J. T., Mourao-Miranada, J., Kather, J. N., Kording, K., Richards, B., & Bzdok, D. (2020). Different scaling of linear models and deep learning in ukbiobank brain images versus machine-learning datasets. *Nature Communications*, *11*(1), 4238. https://doi.org/10.1038/s41467-020-18037-z

Scott, K. (2020, September 22). Microsoft teams up with openai to exclusively license gpt-3 language model. https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python, 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Segal, A., Parkes, L., Aquino, K., Kia, S. M., Wolfers, T., Franke, B., Hoogman, M., Beckmann, C. F., Westlye, L. T., Andreassen, O. A., Zalesky, A., Harrison, B. J., Davey, C. G., Soriano-Mas, C., Cardoner, N., Tiego, J., Yücel, M., Braganza, L., Suo, C., … Fornito, A. (2023). Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. *Nature Neuroscience*, 1–17. https://doi.org/10.1038/s41593-023-01404-6

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in mri. *NeuroImage*, *22*(3), 1060–1075. https://doi.org/10.1016/j.neuroimage.2004.03.032

Shahab, S., Mulsant, B. H., Levesque, M. L., Calarco, N., Nazeri, A., Wheeler, A. L., Foussias, G., Rajji, T. K., & Voineskos, A. N. (2019). Brain structure, cognition, and brain age in schizophrenia, bipolar disorder, and healthy controls. *Neuropsychopharmacology*, *44*(5), 898–906. https://doi.org/10.1038/s41386-018-0298-z

Sharma, R., Goel, T., Tanveer, M., Lin, C. T., & Murugan, R. (2023). Deep-learning-based diagnosis and prognosis of alzheimer's disease: A comprehensive review. *IEEE Transactions on Cognitive and Developmental Systems*, *15*(3), 1123–1138. https://doi.org/10.1109/TCDS.2023.3254209

Shmueli, G. (2010). To explain or to predict?

Shoeibi, A., Khodatars, M., Jafari, M., Moridian, P., Rezaei, M., Alizadehsani, R., Khozeimeh, F., Gorriz, J. M., Heras, J., Panahiazar, M., Nahavandi, S., & Acharya, U. R. (2021). Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review. *Computers in Biology and Medicine*, *136*, 104697. https://doi.org/10.1016/j.compbiomed.2021.104697

Siegert, R. J., & Abernethy, D. A. (2005). Depression in multiple sclerosis: A review. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(4), 469–475. https://doi.org/10.1136/jnnp.2004.054635

Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. https://doi.org/10.48550/arXiv.1312.6034

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556 [cs]*. http://arxiv.org/abs/1409.1556

Sixt, L., Granz, M., & Landgraf, T. (2020). When explanations lie: Why many modified bp attributions fail. *arXiv:1912.09818 [cs, stat]*. http://arxiv.org/abs/1912.09818

Smith, L. N. (2017). Cyclical learning rates for training neural networks. *arXiv:1506.01186 [cs]*. http://arxiv.org/abs/1506.01186

Smith, S. M., Elliott, L. T., Alfaro-Almagro, F., McCarthy, P., Nichols, T. E., Douaud, G., & Miller, K. L. (2020). Brain aging comprises many modes

of structural and functional change with distinct genetic and biophysical associations (J. E. Peelle, F. P. de Lange, C. Madan, & L. Nyberg, Eds.). *eLife*, *9*, e52677. https://doi.org/10.7554/eLife.52677

Smith, S. M., & Nichols, T. E. (2018). Statistical challenges in "big data" human neuroimaging. *Neuron*, *97*(2), 263–268. https://doi.org/10.1016/j.neuron.2017.12.018

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E., & Miller, K. L. (2019). Estimation of brain age delta from brain imaging. *NeuroImage*, *200*, 528–539. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.06.017

Smolensky, P. (1987). Connectionist ai, symbolic ai, and the brain. *Artificial Intelligence Review*, *1*(2), 95–109. https://doi.org/10.1007/BF00130011

Sone, D., Beheshti, I., Shinagawa, S., Niimura, H., Kobayashi, N., Kida, H., Shikimoto, R., Noda, Y., Nakajima, S., Bun, S., Mimura, M., & Shigeta, M. (2022). Neuroimaging-derived brain age is associated with life satisfaction in cognitively unimpaired elderly: A community-based study. *Translational Psychiatry*, *12*(1), 1–6. https://doi.org/10.1038/s41398-022-01793-5

Sporns, O. (2013). Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, *15*(3), 247–262.

Squires, M., Tao, X., Elangovan, S., Gururajan, R., Zhou, X., Li, Y., & Acharya, U. R. (2023). Identifying predictive biomarkers for repetitive transcranial magnetic stimulation response in depression patients with explainability. *Computer Methods and Programs in Biomedicine*, *242*, 107771. https://doi.org/10.1016/j.cmpb.2023.107771

Srećković, S., Berber, A., & Filipović, N. (2022). The automated laplacean demon: How ML challenges our views on prediction and explanation. *Minds and Machines*, *32*(1). https://doi.org/10.1007/s11023-021-09575-6

Steen, R. G., Hamer, R. M., & Lieberman, J. A. (2007). Measuring brain volume by mr imaging: Impact of measurement precision and natural variation on sample size requirements. *AJNR: American Journal of Neuroradiology*, *28*(6), 1119. https://doi.org/10.3174/ajnr.A0537

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). Uk biobank: An open access resource for identifying the causes of

a wide range of complex diseases of middle and old age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Symms, M., Jäger, H. R., Schmierer, K., & Yousry, T. A. (2004). A review of structural magnetic resonance neuroimaging. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(9), 1235–1244. https://doi.org/10.1136/jnnp.2003.032714

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1

Tanveer, M., Ganaie, M. A., Beheshti, I., Goel, T., Ahmad, N., Lai, K.-T., Huang, K., Zhang, Y.-D., Del Ser, J., & Lin, C.-T. (2023). Deep learning for brain age estimation: A systematic review. *Information Fusion*, *96*, 130–143. https://doi.org/10.1016/j.inffus.2023.03.007

Tau, G. Z., & Peterson, B. S. (2010). Normal development of brain circuits. *Neuropsychopharmacology*, *35*(1), 147–168. https://doi.org/10.1038/npp.2009.115

Teipel, S., Drzezga, A., Grothe, M. J., Barthel, H., Chételat, G., Schuff, N., Skudlarski, P., Cavedo, E., Frisoni, G. B., Hoffmann, W., Thyrian, J. R., Fox, C., Minoshima, S., Sabri, O., & Fellgiebel, A. (2015). Multimodal imaging in alzheimer's disease: Validity and usefulness for early detection. *The Lancet Neurology*, *14*(10), 1037–1053. https://doi.org/10.1016/S1474-4422(15)00093-9

the GTEx Consortium. (2015). Human genomics. the genotype-tissue expression (gtex) pilot analysis: Multitissue gene regulation in humans. *Science (New York, N.Y.)*, *348*(6235), 648–660. https://doi.org/10.1126/science.1262110

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., … Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the mcdonald criteria.

The Lancet Neurology, *17*(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolín, S., Bralten, J., Bruin, W. B., Bülow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyler, L. T., Faraone, S. V., Favre, P., … Zelman, V. (2020). Enigma and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry*, *10*(1), 1–28. https://doi.org/10.1038/s41398-020-0705-1

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatrics Society*, *40*(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x

Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. *Proceedings of the 4th Machine Learning for Healthcare Conference*, 359–380. Retrieved December 1, 2023, from https://proceedings.mlr.press/v106/tonekaboni19a.html

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., … Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. http://arxiv.org/abs/2307.09288

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C.-Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., … O'Donovan, M. C. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, *604*(7906), 502–508. https://doi.org/10.1038/s41586-022-04434-5

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, *33*(1), 1–13. https://doi.org/10.1002/hbm.21186

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 1–21. https://doi.org/10.1038/s43586-021-00056-9

Upthegrove, R., Marwaha, S., & Birchwood, M. (2017). Depression and schizophrenia: Cause, consequence, or trans-diagnostic issue? *Schizophrenia Bulletin*, *43*(2), 240–244. https://doi.org/10.1093/schbul/sbw097

Valdes-Hernandez, P. A., Laffitte Nodarse, C., Peraza, J. A., Cole, J. H., & Cruz-Almeida, Y. (2023). Toward mr protocol-agnostic, unbiased brain age predicted from clinical-grade mris. *Scientific Reports*, *13*(1), 19570. https://doi.org/10.1038/s41598-023-47021-y

VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M., & Kraft, P. (2014). Methodological challenges in mendelian randomization. *Epidemiology (Cambridge, Mass.)*, *25*(3), 427–435. https://doi.org/10.1097/EDE.0000000000000081

van Erp, T. G. M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., Pearlson, G. D., Yao, N., Fukunaga, M., Hashimoto, R., Okada, N., Yamamori, H., Bustillo, J. R., Clark, V. P., Agartz, I., Mueller, B. A., Cahn, W., de Zwarte, S. M. C., Hulshoff Pol, H. E., … Turner, J. A. (2018). Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the enhancing neuro imaging genetics through meta analysis (enigma) consortium. *Biological Psychiatry*, *84*(9), 644–654. https://doi.org/10.1016/j.biopsych.2018.04.023

Varoquaux, G., & Colliot, O. (2023). Evaluating machine learning models and their diagnostic value. In O. Colliot (Ed.), *Machine learning for brain disorders* (pp. 601–630). Springer US. https://doi.org/10.1007/978-1-0716-3195-9_20

Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: Methodological failures and recommendations for the future. *npj Digital Medicine*, *5*(1), 1–8. https://doi.org/10.1038/s41746-022-00592-y

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Venkatraghavan, V., Voort, v. d. S. R., Bos, D., Smits, M., Barkhof, F., Niessen, W. J., Klein, S., & Bron, E. E. (2023). Computer-aided diagnosis and prediction in brain disorders. In O. Colliot (Ed.), *Machine learning for*

brain disorders (pp. 459–490). Springer US. https://doi.org/10.1007/978-1-0716-3195-9_15

Verbanck, M., Chen, C.-Y., Neale, B., & Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature Genetics*, *50*(5), 693–698. https://doi.org/10.1038/s41588-018-0099-7

Verdi, S., Marquand, A. F., Schott, J. M., & Cole, J. H. (2021). Beyond the average patient: How neuroimaging models can address heterogeneity in dementia. *Brain*, *144*(10), 2946–2953. https://doi.org/10.1093/brain/awab165

Verma, S., Goel, T., Tanveer, M., Ding, W., Sharma, R., & Murugan, R. (2023). Machine learning techniques for the schizophrenia diagnosis: A comprehensive review and future research directions. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 4795–4807. https://doi.org/10.1007/s12652-023-04536-6

Vidal-Pineiro, D., Wang, Y., Krogsrud, S. K., Amlien, I. K., Baaré, W. F., Bartres-Faz, D., Bertram, L., Brandmaier, A. M., Drevon, C. A., Düzel, S., Ebmeier, K., Henson, R. N., Junqué, C., Kievit, R. A., Kühn, S., Leonardsen, E., Lindenberger, U., Madsen, K. S., Magnussen, F., … Fjell, A. (2021). Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change (J. Zhou, C. Büchel, & X.-N. Zuo, Eds.). *eLife*, *10*, e69995. https://doi.org/10.7554/eLife.69995

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Contributors, S. 1. (2020). Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vogt, N. (2023). Reproducibility in mri. *Nature Methods*, *20*(1), 34–34. https://doi.org/10.1038/s41592-022-01737-3

Wain, L. V., Shrine, N., Miller, S., Jackson, V. E., Ntalla, I., Soler Artigas, M., Billington, C. K., Kheirallah, A. K., Allen, R., Cook, J. P., Probert, K., Obeidat, M., Bossé, Y., Hao, K., Postma, D. S., Paré, P. D., Ramasamy, A., (UKBEC), U. B. E. C., Mägi, R., … Hall, I. P. (2015). Novel insights into the genetics of smoking behaviour, lung function,

and chronic obstructive pulmonary disease (uk bileve): A genetic association study in uk biobank. *The Lancet. Respiratory Medicine*, *3*(10), 769–781. https://doi.org/10.1016/S2213-2600(15)00283-0

Wang, D., Honnorat, N., Fox, P. T., Ritter, K., Eickhoff, S. B., Seshadri, S., & Habes, M. (2023). Deep neural network heatmaps capture alzheimer's disease patterns reported in a large meta-analysis of neuroimaging studies. *NeuroImage*, *269*, 119929. https://doi.org/10.1016/j.neuroimage.2023.119929

Wang, J., Knol, M. J., Tiulpin, A., Dubost, F., Bruijne, d. M., Vernooij, M. W., Adams, H. H. H., Ikram, M. A., Niessen, W. J., & Roshchupkin, G. V. (2019). Gray matter age prediction as a biomarker for risk of dementia. *Proceedings of the National Academy of Sciences*, *116*(42), 21213–21218. https://doi.org/10.1073/pnas.1902376116

Wang, R., Erus, G., Chaudhari, P., & Davatzikos, C. (2023). Adapting machine learning diagnostic models to new populations using a small amount of data: Results from clinical neuroscience. https://doi.org/10.48550/arXiv.2308.03175

Wang, W. Y. S., Barratt, B. J., Clayton, D. G., & Todd, J. A. (2005). Genome-wide association studies: Theoretical and practical concerns. *Nature Reviews. Genetics*, *6*(2), 109–118. https://doi.org/10.1038/nrg1522

Wang, Y., Gong, N., & Fu, C. (2021). Major depression disorder diagnosis and analysis based on structural magnetic resonance imaging and deep learning. *Journal of Integrative Neuroscience*, *20*(4), 977–984. https://doi.org/10.31083/j.jin2004098

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with fuma. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Siuciak, J. A., Soares, H., Toga, A. W., Trojanowski, J. Q., & Initiative, A. D. N. (2012). The alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *8*(1 Suppl), S1–68. https://doi.org/10.1016/j.jalz.2011.09.172

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Morris, J. C., Petersen, R. C., Salazar, J., Saykin, A. J., Shaw, L. M., Toga, A. W., & Trojanowski, J. Q. (2017). The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *13*(5), 561–571. https://doi.org/10.1016/j.jalz.2016.10.006

Werkhoven, S., Anderson, J. H., & Robeyns, I. A. M. (2022). Who benefits from diagnostic labels for developmental disorders? *Developmental Medicine & Child Neurology*, *64*(8), 944–949. https://doi.org/10.1111/dmcn.15177

Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., Brown, S. M., Pavel, M., Erdogmus, D., Brooks, D. H., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences*, *27*(3), 246–257. https://doi.org/10.1016/j.tics.2022.12.015

Whelan, R., & Garavan, H. (2014). When optimism hurts: Inflated predictions in psychiatric neuroimaging. *Biological Psychiatry*, *75*(9), 746–748. https://doi.org/10.1016/j.biopsych.2013.05.014

Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., & Marquand, A. F. (2015). From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience & Biobehavioral Reviews*, *57*, 328–349. https://doi.org/10.1016/j.neubiorev.2015.08.001

Wolfers, T., Doan, N. T., Kaufmann, T., Alnæs, D., Moberget, T., Agartz, I., Buitelaar, J. K., Ueland, T., Melle, I., Franke, B., Andreassen, O. A., Beckmann, C. F., Westlye, L. T., & Marquand, A. F. (2018). Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry*, *75*(11), 1146–1155. https://doi.org/10.1001/jamapsychiatry.2018.2467

Wolfers, T., Rokicki, J., Alnæs, D., Berthet, P., Agartz, I., Kia, S. M., Kaufmann, T., Zabihi, M., Moberget, T., Melle, I., Beckmann, C. F., Andreassen, O. A., Marquand, A. F., & Westlye, L. T. (2021). Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *Human Brain Mapping*, *42*(8), 2546–2555. https://doi.org/10.1002/hbm.25386

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Wood, D. A., Kafiabadi, S., Busaidi, A. A., Guilhem, E., Montvila, A., Lynch, J., Townend, M., Agarwal, S., Mazumder, A., Barker, G. J., Ourselin, S., Cole, J. H., & Booth, T. C. (2022). Accurate brain-age models for routine clinical mri examinations. *NeuroImage*, *249*, 118871. https://doi.org/10.1016/j.neuroimage.2022.118871

World Health Organization. (2001). The world health report 2001: Mental health: New understanding, new hope.

World Health Organization. (2021). Global status report on the public health response to dementia.

World Health Organization. (2022). World mental health report: Transforming mental health for all.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., … Sullivan, P. F. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yin, C., Imms, P., Cheng, M., Amgalan, A., Chowdhury, N. F., Massett, R. J., Chaudhari, N. N., Chen, X., Thompson, P. M., Bogdan, P., Irimia, A., & the Alzheimer's Disease Neuroimaging Initiative. (2023). Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment. *Proceedings of the National Academy of Sciences*, *120*(2), e2214634120. https://doi.org/10.1073/pnas.2214634120

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv:1506.06579 [cs]*. http://arxiv.org/abs/1506.06579

Young, A. L., Marinescu, R. V., Oxtoby, N. P., Bocchetta, M., Yong, K., Firth, N. C., Cash, D. M., Thomas, D. L., Dick, K. M., Cardoso, J., van

Swieten, J., Borroni, B., Galimberti, D., Masellis, M., Tartaglia, M. C., Rowe, J. B., Graff, C., Tagliavini, F., Frisoni, G. B., … Alexander, D. C. (2018). Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nature Communications*, *9*(1), 4273. https://doi.org/10.1038/s41467-018-05892-0

Zabihi, M., Kia, S. M., Wolfers, T., Dinga, R., Arenas, A. L., Bzdok, D., Beckmann, C. F., & Marquand, A. (2021, March 14). *Non-linearity matters: A deep learning solution to generalization of hidden brain patterns across population cohorts* (tech. rep.). Neuroscience. http://biorxiv.org/lookup/doi/10.1101/2021.03.10.434856

Zeiler, M. D., & Fergus, R. (2013). Visualizing and understanding convolutional networks. http://arxiv.org/abs/1311.2901

Zeune, L. L., Boink, Y. E., van Dalum, G., Nanou, A., de Wit, S., Andree, K. C., Swennenhuis, J. F., van Gils, S. A., Terstappen, L. W. M. M., & Brune, C. (2020). Deep learning of circulating tumour cells. *Nature Machine Intelligence*, *2*(2), 124–133. https://doi.org/10.1038/s42256-020-0153-x

Zhang, L., Wang, M., Liu, M., & Zhang, D. (2020). A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in Neuroscience*, *14*. https://www.frontiersin.org/article/10.3389/fnins.2020.00779

Zhao, Q., Wang, J., Hemani, G., Bowden, J., & Small, D. S. (2020). Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *The Annals of Statistics*, *48*(3), 1742–1769. https://doi.org/10.1214/19-AOS1866

Zheng, A., & Casari, A. (2018, March 23). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2015). Object detectors emerge in deep scene cnns. http://arxiv.org/abs/1412.6856

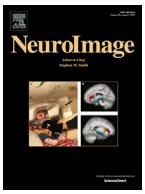
I

NeuroImage 256 (2022) 119210

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



Deep neural networks learn general and clinically relevant representations of the ageing brain

Esten H. Leonardsen[a,b,*], Han Peng[c], Tobias Kaufmann[b,d], Ingrid Agartz[b,e,f], Ole A. Andreassen[b], Elisabeth Gulowsen Celius[g,h], Thomas Espeseth[a,i], Hanne F. Harbo[g,h], Einar A. Høgestøl[a,b,g], Ann-Marie de Lange[a,j,k], Andre F. Marquand[l], Didac Vidal-Piñeiro[a], James M. Roe[a], Geir Selbæk[m,n], Øystein Sørensen[a], Stephen M. Smith[c], Lars T. Westlye[a,b,o,1], Thomas Wolfers[a,b,1], Yunpeng Wang[a,1]

[a] Department of Psychology, University of Oslo, Oslo, Norway
[b] Norwegian Centre for Mental Disorders Research (NORMENT), Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[c] Wellcome Centre for Integrative Neuroimaging (WIN FMRIB), University of Oxford, Oxford, OX3 9DU, United Kingdom
[d] Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Germany
[e] Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway
[f] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet & Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden
[g] Department of Neurology, Oslo University Hospital, Norway
[h] Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[i] Department of Psychology, Bjørknes University College, Oslo, Norway
[j] LREN, Centre for Research in Neurosciences-Department of Clinical Neurosciences, CHUV and University of Lausanne, Lausanne, Switzerland
[k] Department of Psychiatry, University of Oxford, Oxford, UK
[l] Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, Netherlands
[m] Norwegian National Advisory Unit on Aging and Health, Vestfold Hospital Trust, Tønsberg, Norway
[n] Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway
[o] KG Jebsen Center for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway

A B S T R A C T

The discrepancy between chronological age and the apparent age of the brain based on neuroimaging data — the brain age delta — has emerged as a reliable marker of brain health. With an increasing wealth of data, approaches to tackle heterogeneity in data acquisition are vital. To this end, we compiled raw structural magnetic resonance images into one of the largest and most diverse datasets assembled (n=53542), and trained convolutional neural networks (CNNs) to predict age. We achieved state-of-the-art performance on unseen data from unknown scanners (n=2553), and showed that higher brain age delta is associated with diabetes, alcohol intake and smoking. Using transfer learning, the intermediate representations learned by our model complemented and partly outperformed brain age delta in predicting common brain disorders. Our work shows we can achieve generalizable and biologically plausible brain age predictions using CNNs trained on heterogeneous datasets, and transfer them to clinical use cases.

Introduction

Neurodevelopmental and age-related changes in the brain play a crucial role in the etiology of complex neurological Stephan and Brayne (2009) and mental disorders (Thapar and Riglin, 2020). Predictive models for an individual's age based on magnetic resonance imaging (MRI) have been used to estimate normative trajectories across the lifespan (Cole and Franke, 2017; Cole et al., 2017; Franke et al., 2012; 2010). Individual deviations from these trajectories, often called the brain age delta, have been linked to brain health (Franke and Gaser, 2019; Kaufmann et al., 2019), and more extreme deviations observed in patients with schizophrenia (SCZ) (Nenadi et al., 2017; Rokicki et al., 2021; Schnack et al., 2016), depression (Han et al., 2020), cognitive impairment (Elliott et al., 2019; Liem et al., 2017), dementia (Wang et al., 2019), Alzheimer's disease (AD) (Gaser et al., 2013) and multiple sclerosis (MS) (Høgestøl et al., 2019), implying that such deviations could be a feasible biological marker for various brain disorders.

The brain age of an individual is typically estimated from brain images using statistical learning techniques. The first-generation models were relatively simple, typically based on independent voxels

* Corresponding author. Postboks 1094 Blindern, 0317 OSLO.
E-mail address: estenhl@psykologi.uio.no (E.H. Leonardsen).
[1] equal contribution

https://doi.org/10.1016/j.neuroimage.2022.119210.
Received 23 December 2021; Received in revised form 16 March 2022; Accepted 11 April 2022
Available online 21 April 2022.

Franke et al. (2010) or a limited number of imaging-derived phenotypes (IDPs) reflecting brain properties such as volumetric measures of different regions (Smith et al., 2019). These models generally estimated linear relationships, were restricted in scale, and were trained on datasets with tens, hundreds or a few thousand participants (Franke and Gaser, 2019). In parallel with continuous computational advances the exponential growth of MRI data has enabled deep learning models of scale for accurate brain age estimation (Abrol et al., 2021; Cole et al., 2017). Deep learning models can take minimally- or non-preprocessed 3-D images as input - avoiding computationally demanding and hypothesis-driven (Oquendo et al., 2012) choices during image processing - and model complex non-linear relations between voxels. One such model is the Simple Fully Convolutional Network (SFCN), a novel deep Convolutional Neural Network (CNN) that won the Predictive Analysis Challenge for brain age prediction in 2019 (PAC2019) (Gong et al., 2021; Peng et al., 2021). While such deep learning approaches allow for the prediction of brain age with unprecedented accuracy, and can potentially help us identify idiosyncratic regional patterns of neurodevelopment and ageing at the individual level (Dinsdale et al., 2021; Hofmann et al., 2021), their complexity also comes with a risk of overfitting, namely finding patterns in the training data which do not generalize well to new, previously unseen, participants (Arbabshirani et al., 2017).

Transfer learning, a deep learning technique widely used in other areas of applied machine learning research, has recently gained momentum in neuroimaging (Valverde et al., 2021). Here, learned intermediate representations can be shared between tasks, allowing a model to be transferred to a problem or a dataset it was initially not trained for (Bengio, 2012). This approach has arguably been one of the core developments underlying the practical success of deep learning in a range of computer vision problems (SharifRazavian et al., 2014) by using models trained on general purpose datasets, typically ImageNet (Russakovsky et al., 2015), and fine-tuning them towards a wide array of tasks. Recent studies have shown that transfer learning yields promising results also for brain age predictions (Jonsson et al., 2019) and clinical classifications based on MRI data (Bashyam et al., 2020; Lu et al., 2020). This exemplifies the need for robust, pretrained models on massive multisite datasets, that can be translated to smaller clinical samples, and, ultimately, to individual cases in a clinical setting. Modelling brain age as a pre-training step has obvious advantages as age is a variable which is available in most current MRI datasets. Additionally, the representations learned by the brain age models, representing partly independent dimensions of age-related variance, could be of direct importance in individual level brain phenotyping.

In the present study we trained deep neural network models for brain age prediction on structural MRI data from 53,542 healthy individuals between 3 and 95 years of age to test their ability to generalize, and demonstrate the downstream applicability and biological relevance of a properly generalizing model. We used the Simple Fully Convolutional Network with a softmax output (SFCN-sm) which predicts age as a discrete probability distribution, and compared its accuracy and generalizability with two proposed variants of the architecture; a regression variant (SFCN-reg), directly predicting continuous age, and a ranking variant (SFCN-rank) encoding age as an ordinal vector. Brain age delta was computed as the difference between the predicted and chronological age. To estimate the sensitivity and clinical relevance of our best model, we tested for associations between brain age delta and a range of clinical and biological phenotypes, and sociodemographic and lifestyle variables, in unseen data from a population sample. To further demonstrate the applicability of the model, we employed the pretrained SFCN-reg in a transfer learning context to predict case-control status for AD, MS, mild cognitive impairment (MCI), SCZ, mood disorders and psychotic disorders on datasets obtained from a range of MRI scanners. To promote transparency and reproducibility we have implemented an easy-to-use Keras interface for all the trained models, both the brain age and clinical predictors, and a pipeline for preprocessing images, available on our GitHub http://www.github.com/estenhl/pyment-public.

Results

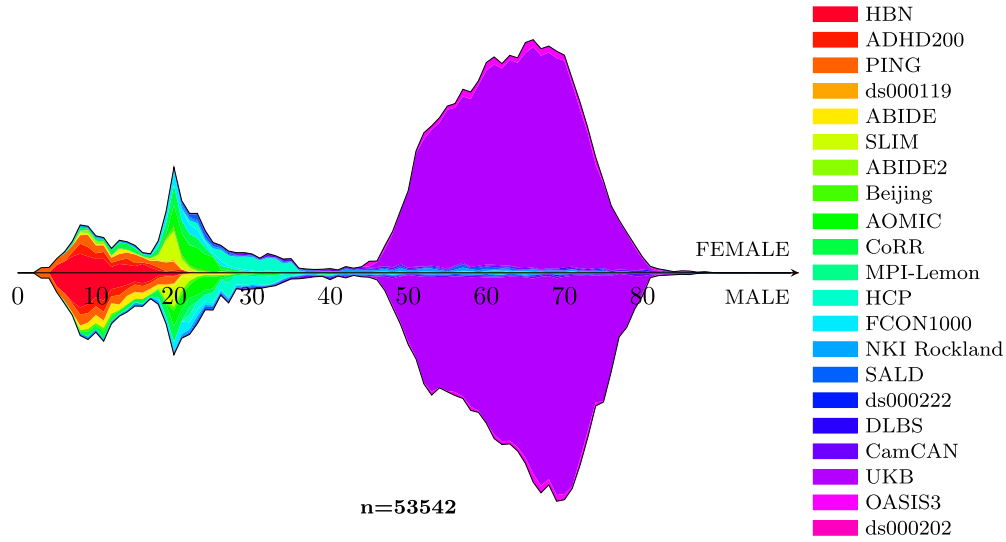
We compiled 21 publicly available datasets with T1-weighted MRI scans into a large and diverse imaging dataset (total N = 53542; female N = 27715; age range = 3-95), and trained a Simple Fully Convolutional 3-dimensional CNN with 6 convolutional blocks and a softmax output layer (SFCN-sm, Fig. 1b), as introduced in Peng et al (Peng et al., 2021). We then proposed two alternatives for the prediction layer of this architecture, the first based on regression (SFCN-reg) and the second based on ranking (Chen et al., 2017) (SFCN-rank) (see Materials and Methods for details). Due to the time-consuming process of model evaluation, we restricted our search to these three variants of the given architecture, and trained a handful of versions of each variant with different hyperparameter settings. To evaluate the generalization performance of the models, we divided our data at two levels; A reference dataset (Fig. 1 and Supplementary Table 1) and an external dataset (Fig. 2a and Supplementary Table 2). In the reference dataset, we evaluated the performance of the trained models on an independent test split from known scanners with an age distribution resembling the training split. We then tested model performance on the external dataset compiled from different sources, originating from scanners unseen by the models during training with a divergent age distribution.

Superior generalization performance of SFCN-reg

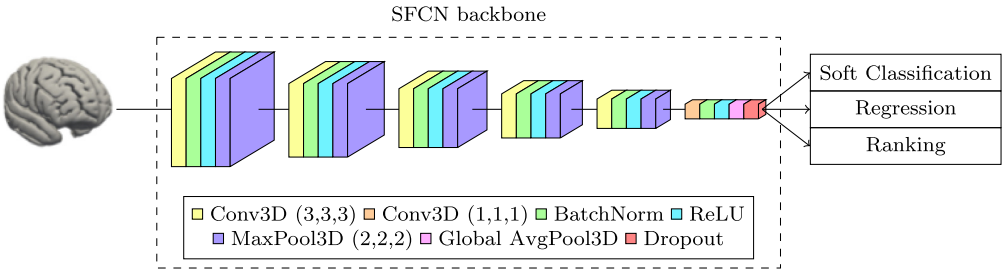
We used the training and validation sets to optimize and tune the models, and a conjunction of the test set and the external dataset in a final model comparison. For each model variant we trained three versions, with different hyperparameter settings, on the training data, and selected the best model based on the mean absolute error (MAE) on the validation set (Fig. 1c and Supplementary Table 8). We then compared the performances of the best version for each model variant on the test set. In this comparison the results mirrored those of the validation set, with SCFN-sm achieving the best result with an MAE of 2.23 years, followed by SCFN-reg with 2.47 years and SCFN-rank with 2.55 years (Fig. 2b). Given the added complexity of including multiple datasets from a large range of scanners, we consider these results to be approximately on par with the MAE of 2.14 reported in the original SFCN paper (Peng et al., 2021), and thus among the best performing models in the field He et al. (2021). Additionally, the heterogeneous origins of the dataset facilitate cross-site generalization, an essential property when training large multisite models (Dockès et al., 2021).

As a conclusive test of model generalization, we performed the same comparison on the external dataset, containing unseen data from different MRI scanners with an age distribution diverging from that of the reference dataset. SCFN-reg substantially outperformed the two alternatives, with an MAE of 3.90 compared to 5.04 and 5.82 for SFCN-sm and SFCN-rank respectively (Fig. 2). While the performance of all models were lower on the external dataset, the extent of the generalization error was considerably different. When compared with MAEs from the test set, the average error of SFCN-reg increased by approximately half, while it more than doubled in SFCN-sm and SFCN-rank (Supplementary Table 9). This difference coincides with the architectural differences between the models: Where both the SFCN-sm and SFCN-rank used age-bins, with an output node for each age in its prediction range, SFCN-reg had a single output predicting a single continuous number. Therefore, the predictions of the SFCN-reg reflect a simpler combination of the learned representations in the preceding layer of the model, which we hypothesize may be the reason for the improved generalization performance. In the subsequent applications we use the SFCN-reg version that achieved the best MAE on the external dataset. Additionally, to facilitate cross-study comparisons, we have compiled a range of performance metrics for our three models on the external dataset in Table 1.

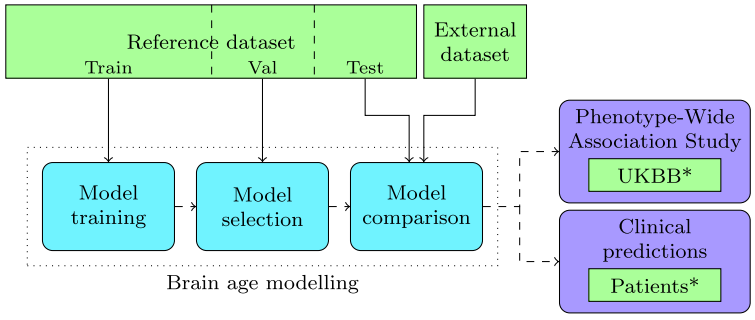
To better understand the visual patterns underlying the predictions of the SFCN-reg we conducted two post-hoc correlational analyses. First,



(a) Age distribution of the reference dataset



(b) Brain age model architecture

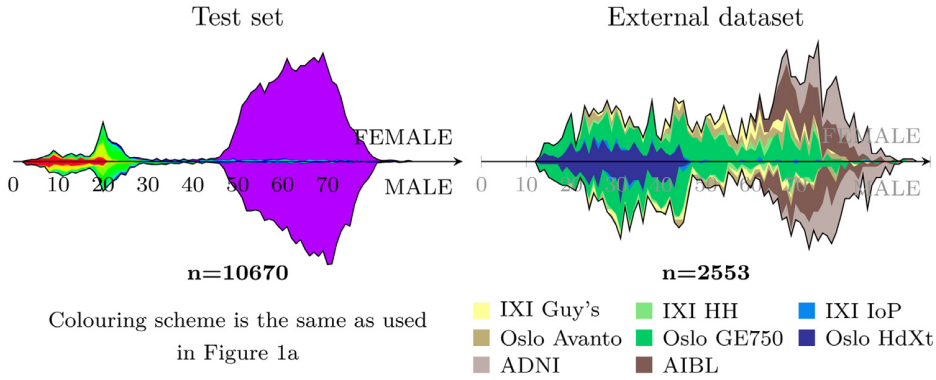


(c) Flowchart of the modelling process and applications

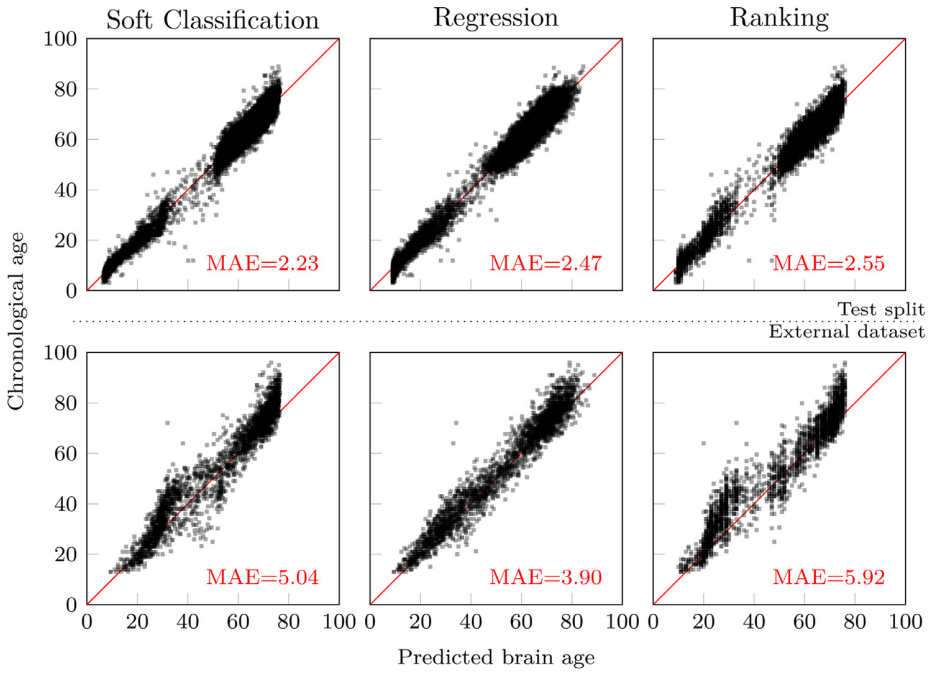
Fig. 1. An overview of the dataset and models used for brain age modelling. (a) The reference dataset contains 53,542 healthy participants from 21 datasets, with ages ranging from 3 to 95 years. (b) We implemented three model variants for predicting brain age, all based on the contest-winning SFCN architecture (Peng et al., 2021). All models take minimally preprocessed T1-weighted MRI images as input. (c) The modelling process consisted of three steps, utilizing different parts of the reference data and the external dataset from previously unseen scanners. The best brain age model was applied in a phenome-wide association study, and a case-control comparison including several clinical conditions.

we examined brains in groups of participants with extreme brain age deltas. Second, we correlated the delta with FreeSurfer-extracted imaging measures. High deltas were broadly associated with a general pattern of reduced cortical thickness (Fig. 3a and Supplementary Figure 8), reduced volumes of several subcortical areas, and increases in ventricle size, cerebrospinal fluid, and both white matter and non-white matter hypointensities (Fig. 3b). We also correlated the delta with voxel-wise volume and area, seeing a less pronounced pattern (Supplementary Figure 9). A full overview of the subcortical correlations can be seen in Supplementary Table 11.

Observing that all the models performed worse in the external dataset than in the test set, we performed post-hoc analyses to further understand the causes underlying the generalization problems. First, we investigated whether the inflated error could be explained by an additive offset, a linear multiplicative bias, or a general increase in variance. We observed that all three measures were amplified in the exter-



(a) Age distribution of the two datasets used in the comparison



(b) Brain age predictions

Fig. 2. The two datasets used in the model comparison, and the predictive performance of the three model variants. (a) For comparing the models we employed two distinct datasets: A test set sampled from the reference dataset, and an external dataset. The former was drawn from the reference dataset using stratification, and as a consequence has a similar age and sex-distribution. The latter was compiled from a different subset of datasets, and thus was acquired from different scanners. The age range of the external dataset is somewhat narrower, spanning a region of 13 to 95 years, and has a more uniform distribution. (b) The six scatter plots display the predictions of a given model on the x-axis, against the ground truth age of the participants on the y-axis. The top row contains the predictions of the three model variants on the test set, and the bottom row the external dataset.

Table 1
Predictive performance of the models. We compared the soft classifiation model (SFCN-sm), the regression model (SFCN-reg) and the ranking model (SFCN-rank) on the external dataset, originating from scanners which has not been seen by the models during training. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R and R2 are sensitive to the age range of the dataset, while normalized RMSE (nRMSE) and Relative Absolute Error (RAE) are not, facilitating comparisons across datasets.

Model	MAE	RMSE	R	R^2	mRMSE	RAE
SFCN-sm	5.04	6.51	0.961	0.903	0.078	0.26
SFCN-reg	**3.90**	**5.11**	**0.975**	**0.940**	**0.061**	**0.20**
SFCN-rank	5.92	7.54	0.959	0.870	0.090	0.31

nal dataset when compared to the test set, and thus jointly contributed to the increase (Supplementary Table 10). Secondly, we tried isolating two sources of generalization error: Differences in population, represented here by age and sex distributions, and differences in scanners and acquisition protocols. We approached this by resampling two artificial datasets, both with participants previously unseen by the models. First, we sampled a dataset with an "Unknown population", with participants from the test set following the empirical age and sex distribution of the external dataset. Secondly, we created a dataset with "Unknown scanners", sampling participants from the external dataset while following the distributions of the test set (see Materials and Methods for further details). Due to the stratification used in the initial train/validation/test split the latter set also directly matches the distributions of the training set. For each of these two new datasets we computed an MAE per model, which naturally fell between the MAEs on the test set and on the external dataset. While this approach is exploratory and inherently limited to the characteristics and actual data points making up our datasets, the results clearly indicate that the main driver of generalization error is the unknown scanners (Table 2 and Supplementary Figure 1) which had higher errors (MAE_{US}) than the unknown population (MAE_{UP}) for all the models. Additionally, these two sources of generalization error seem to work additively, with their sum closely matching the full gen-

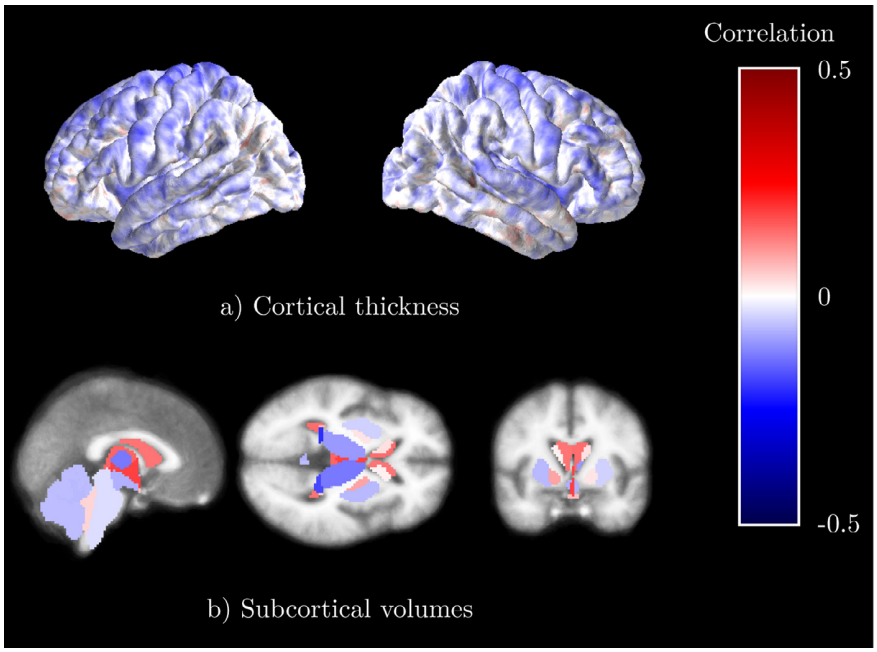


Fig. 3. Correlation between brain age delta and vertex-wise and subcortical measures in imaging space. (a) Correlations with cortical thickness at the vertex-level computed by FreeSurfer, plotted on two generic surface hemispheres. (b) Correlations with subcortical volumes using the standard FreeSurfer atlas, overlayed over an average brain. The reported associations were computed as Pearson correlations in the test portion of UKBB, with the imaging measures derived using FreeSurfer's recon-all pipeline.

Table 2
Results of the post-hoc generalization source analysis. We measured Mean Absolute Errors for the three model variants on the test set, drawn from the same distributions of scanners and ages as the training dataset, the "Unknown population" dataset (MAE_{UP}) and the "Unknown scanners" dataset (MAE_{US}), both representing a single source of generalization error, and the external dataset, different from the training set in both regards.

Model	MAE_{TEST}	MAE_{UP}	MAE_{US}	$MAE_{EXTERNAL}$
SFCN-sm	**2.23**	3.48	3.84	5.04
SFCN-reg	2.47	**2.89**	**3.42**	**3.90**
SFCN-rank	2.55	3.95	4.32	5.92

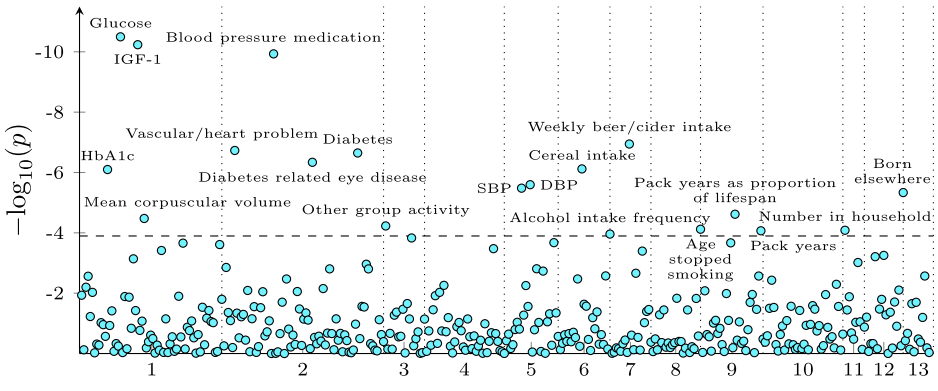
eralization error observed in the external dataset. A final observation was that SFCN-reg handled both sources of error best, which is further evidence of its superior ability to generalize.

Brain age predictions associate with biological phenotypes and lifestyle factors

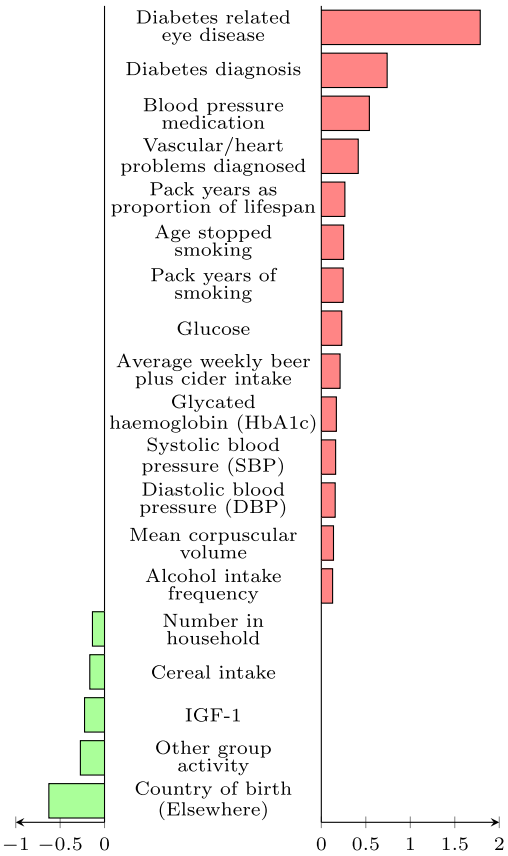
Next, we examined the biological relevance of the model predictions by correlating their deviations from chronological age with an array of phenotypes (n=394) in a phenome-wide association study. We performed this analysis in the subset of the UK Biobank (UKBB) data that was not used for brain age modelling or validation (n=8066), and tested associations with all the biological phenotypes and lifestyle variables accessible, manually divided into thirteen thematic categories for interpretability (Supplementary Table 7). For each phenotype we computed a univariate correlation while correcting for age and sex (Alfaro-Almagro et al., 2021; Smith and Nichols, 2018), and assessed its significance using a Bonferroni-corrected p-value threshold of $p < 1.26 \times 10^{-4}$ (see Materials and Methods). All continuous variables were standardized, such that their effect sizes denote the impact a one standard deviation increase has on the brain age delta. In general, our results corroborated several findings derived from previous studies using smaller samples (Fig. 4a). We observed significantly higher delta in participants with high blood pressure ($\beta = 0.41$, $p = 1.86 \times 10^{-7}$), those currently on blood pressure medication ($\beta = 0.54$, $p = 1.17 \times 10^{-10}$), and a positive correlation with blood pressure readings (diastolic (DBP): $\beta = 0.15$, $p = 2.53 \times 10^{-6}$, systolic (SBP): $\beta = 0.16$, $p = 3.30 \times 10^{-6}$). The associations with the largest effects indicated higher delta in patients with a diabetes diagnosis ($\beta = 0.74$, $p = 2.25 \times 10^{-7}$) or diabetes-related eye problems ($\beta = 1.78$, $p = 4.59 \times 10^{-7}$). Among the biochemical measurements, significant associations with brain age delta were found for blood glucose levels ($\beta = 0.23$, $p = 3.18 \times 10^{-11}$), Insulin-Like Growth Factor-1 levels ($\beta = -0.22$, $p = 5.81 \times 10^{-11}$), glycated haemoglobin levels ($\beta = 0.16$, $p = 7.97 \times 10^{-7}$) and mean corpuscular volume ($\beta = 0.13$, $p = 3.33 \times 10^{-5}$). Associations with variables we categorized as related to diet and lifestyle were dominated by previous smoking, with a positive correlation with number of cigarettes per day (absolute pack years: $\beta = 0.24$, $p = 8.55 \times 10^{-5}$, pack years as proportion of age: $\beta = 0.26$, $p = 2.41 \times 10^{-5}$) and age stopped smoking ($\beta = 0.25$, $p = 7.59 \times 10^{-5}$). We also observed significant associations with average weekly beer and cider intake ($\beta = 0.21$, $p = 1.13 \times 10^{-7}$) and alcohol intake frequency ($\beta = 0.12$, $p = 1.09 \times 10^{-4}$), cereal intake ($\beta = -0.16$, $p = 7.61 \times 10^{-7}$) and participation in "Other group activity" (e.g. social activities not related to a sports or social club, religious group or adult education, $\beta = -0.27$, $p = 5.91 \times 10^{-5}$). Further, we observed a significant correlation with the number of people living in the participants household ($\beta = -0.13$, $p = 8.20 \times 10^{-5}$) and higher deltas in those born outside the United Kingdom and the Republic of Ireland (compared to the baseline group born in England, $\beta = 0.62$, $p = 4.62 \times 10^{-6}$). An overview of all the 394 associations can be found in Supplementary Table 14.

Transferring brain age predictions to developmental and degenerative brain disorders

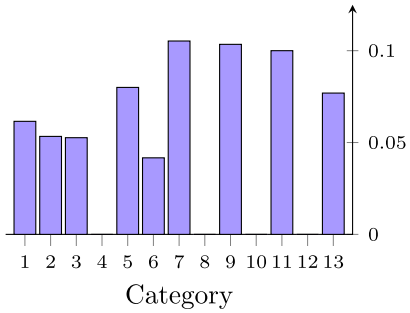
For six different disorders we compiled a patient cohort and a matched control group, and calculated a brain age delta per participant based on the prediction from SFCN-reg (Supplementary Figure 4). In all control groups, the brain age prediction accuracy was approximately the same as for the full test set (MAEs=2.91-4.05, Supplementary Figure 3). Patients with MS showed significantly higher brain age estimates than their matched healthy controls (brain age group mean difference $\Delta = 4.42$ years, $p = 1.71 \times 10^{-22}$, Cohen's $d = 0.87$). A similar pattern was also observed for patients with AD ($\Delta = 2.81$, $p = 4.20 \times 10^{-20}$, $d = 0.59$), MCI ($\Delta = 2.13$, $p = 1.25 \times 10^{-15}$, $d = 0.46$) and SCZ ($\Delta = 1.40$, $p = 4.29 \times 10^{-5}$, $d = 0.34$). For the individuals with mood disorders (MOOD, see Mate-



(a) Significances of the PheWAS correlations



(b) Effect sizes of significant hits



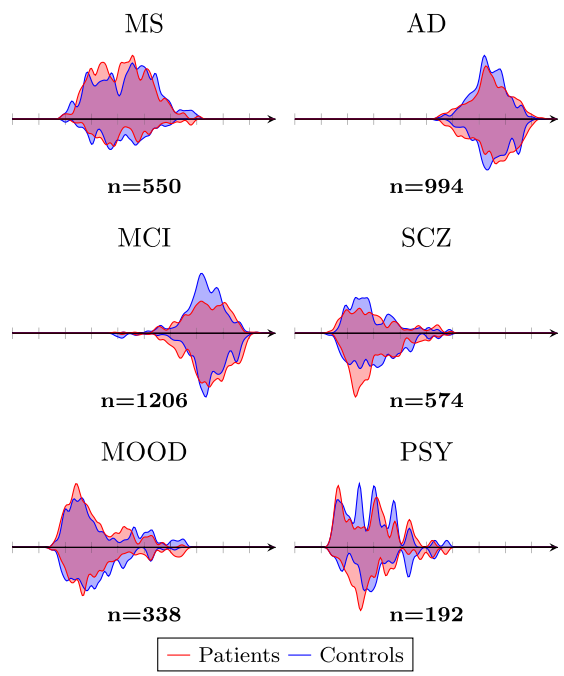
(c) Proportion of significant hits per category

	Category
1	Biochemical measures
2	Health related
3	Lifestyle
4	Cognitive measures
5	Physical measures
6	Diet
7	Alcohol
8	Microbial infection
9	Smoking
10	Mental health
11	Socioeconomic
12	Family
13	Early life factors

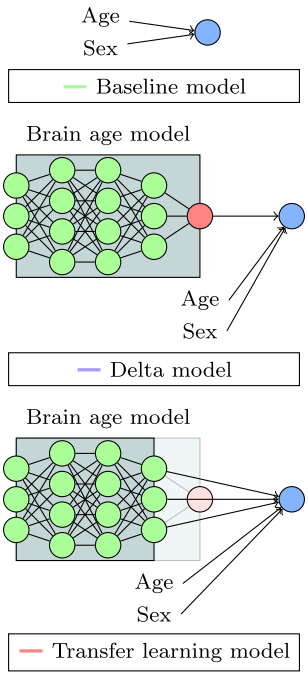
Fig. 4. Associations between the brain age delta and a wide range of phenotypes. We correlated the brain age delta originating from the SFCN-reg with 394 phenotypic variables categorized into thirteen thematic categories. (a) A Manhattan plot visualizing the significances of the 394 associations. A Bonferroni-corrected threshold revealed 19 significant associations with the delta. (b) Effect sizes of the significant associations. For binary variables the effect size express the mean difference between the groups, while for continuous variables it denotes the change in brain age delta associated with a one standard deviation increase. (c) For each of the thirteen categories we calculated the proportion of significant hits by dividing the number of significant hits within that category with the total number of variables in the same category.

rials and Methods) this difference was the smallest ($\Delta = 0.64$, $p = .04$, $d = 0.17$), while the difference was not significant for patients with a mix of psychotic diagnoses (PSY) ($\Delta = 0.74$, $p = .15$, $d = 0.20$). Both the relative ordering of the disorders in terms of group difference, the magnitude of the disparities, and the observed significance resemble a previous study Kaufmann et al. (2019) using a different model based on a smaller dataset.

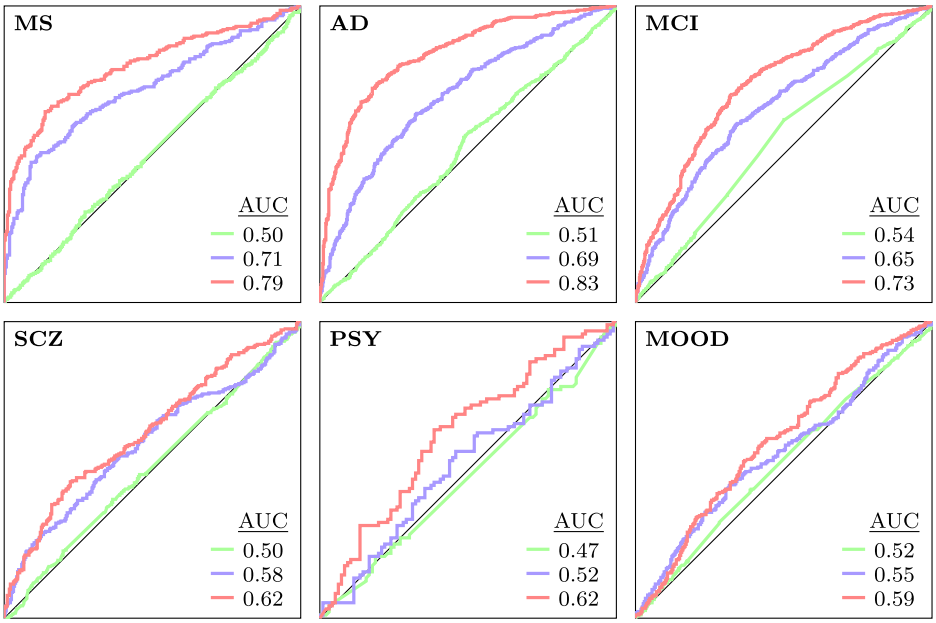
To demonstrate the predictive power of our best performing pretrained brain age model for clinical conditions, we trained multiple binary classifiers to predict whether a participant had a diagnosis or not (Fig. 1c, Fig. 5 and Materials and Methods). We used logistic regression models with an l_1-penalization for this purpose (LASSO models), optimized via a nested cross validation procedure (Supplementary Figure 5), and started with a baseline model classifying participants based only on age and sex. The second model included the brain age delta originating from the brain age prediction of SFCN-reg, and the third model replaced the brain age delta by 64 features encoded in the second to last layer of the same model (Fig. 5b and Materials and Methods). Across the six disorders, the baseline models achieved an area under the receiver operating curve (AUCs) ranging from 0.47 to 0.54, indicating that our matching procedure was satisfactory. Using the second set of models, quantifying the predictive power of the brain age deltas in Supplementary Figure 4, greatly improved the prediction performance when compared to the corresponding baseline models for MS (AUC=0.71 vs.



(a) Patient and control group age distributions



(b) Conceptual model compositions



(c) ROC curves for the predictors

Fig. 5. The datasets and models used for clinical predictions, and an overview over their performance. We trained binary classifiers to separate cases and controls for multiple common brain disorders, using different levels of information from the brain age model. (a) We used a strict matching procedure, drawing a set of controls for each scanner-specific patient dataset matching its empirical age and sex distribution. (b) For each of the six disorders we trained three logistic regression models. The first used age and sex as predictors, the second included the brain age delta, and the third used internal features from SFCN-reg in a transfer learning setting. (c) For each disorder we compared the classifiers using AUCs.

0.50), AD (0.69 vs. 0.51) and MCI (0.65 vs. 0.54); but to a minimal extent for SCZ (0.58 vs. 0.50), MOOD (0.55 vs. 0.52), and PSY (0.52 vs. 0.47). A third set of classifiers were implemented in a strict transfer learning context, utilizing the first part of the SFCN-reg as an encoder. Here, we ran all images in the case-control dataset through the model up until the second-to-last layer, encoding them as 64-dimensional feature vectors. We then used these vectors as predictors in a secondary modelling step, performing the regular cross-validation scheme and training LASSO models with the vectors from the training folds as inputs (Materials and Methods). These features represent high-level, data-driven abstractions of the brain imaging data, and underlie the singular brain age prediction. We refer to this variant of transfer learning as strict because we kept the weights of the initial brain age model locked while optimizing for the new binary objective, which in turn allow us to keep treating these as ageing features and thus promote interpretability. While this complicates contextualizing the performance of our models in terms of existing case-control classifiers, it gives us an indication of the information content of these learned features. This third set of models improved AD prediction substantially (AUC=0.83), and also were notably better for MS (0.79), MCI (0.73) and PSY (0.62), while only a marginal im-

provement was observed in SCZ (0.62) and MOOD (0.59). Overall, our results show that our brain age model can be transferred to make case-control predictions of these common clinical brain disorders.

Discussion

Brain maturation and ageing, and its interactions with clinical brain disorders and conditions, are complex processes with pivotal environmental and genetic contributions (Fjell and Walhovd, 2010; Johnson, 2001). Brain age prediction and the accompanying brain age delta has the potential to provide intuitive and useful measures for summarizing individual brain aberrations. However, technical differences between studies, e.g. the use of different scanners and MRI scan parameters, have represented challenges for the direct generalization and applicability of brain age models based on large training sets. We assembled a large and diverse neuroimaging dataset to train multiple state-of-the-art deep learning models for brain age prediction, and extensively tested their ability to generalize, and their sensitivity to various common brain disorders. Our best model, the SFCN-reg, showed superior performance on an external dataset with differing scanners and age distributions compared to that of the training set, enabling it for applications in other datasets. We then demonstrated the relevance of the model predictions by showing associations between the brain age delta and a range of complex human traits and health outcomes in a population sample. Lastly, we transferred the trained SFCN-reg to clinical data in a transfer learning setting, showing that both the brain age delta and the internal features learned by the model have predictive value when differentiating between controls and patients with common brain disorders.

In our experiment the SFCN-reg outperformed the other models in terms of generalization performance. The age prediction accuracy on the test data (MAE=2.47) is among the best in the field (He et al., 2021). Crucially, the prediction accuracy on data coming from unknown scanners (MAE=3.90) fares very favourably when compared to other studies attempting to transfer between datasets of similar size and complexity (Bashyam et al., 2020; Boyle et al., 2021; He et al., 2021; Jonsson et al., 2019; Ren et al., 2019). Using data from different scanners, protocols and populations in neuroimaging comes with the problem of modelling the effects of these appropriately (Butler et al., 2021; Smith et al., 2019). Previous brain age studies have often explicitly included a scanner term in the modelling or corrected the computed brain age for various biases (Lange A. Marie and Cole, 2020). Recent approaches have tried to address this problem directly via specific deep learning paradigms (Dinsdale et al., 2020). Our results show that given sufficiently large and heterogeneous training data, deep CNNs achieve state-of-the-art performance for brain age predictions even when scanner effects are not explicitly modelled, and more importantly that this performance translates to scanners and protocols that are unknown to the model. This suggests that the representations learned by the model are dominated by age-related variance, not scanner-dependent artefacts, an extension of the model robustness shown in earlier studies (Abrol et al., 2021). From a practical point of view this suggests that our trained model may be employed in other applications in new datasets, without the need for retraining or applying corrective procedures.

Evaluating the biological relevance of the brain age predictions is essential to further understand and trust these models. Therefore, we correlated brain age delta obtained from the SFCN-reg with a wide range of phenotypes in a subset of the UKBB not used for model training. When applying a Bonferroni-corrected threshold, we found nineteen significant hits spread across seven of our thirteen categories. Almost all of these have in earlier studies been found to have a relationship with ageing, either generally or specifically in the brain, or with brain health and/or cognitive function: Glucose level (Karasik et al., 2005), Insulin-like Growth factor-1 (Ashpole et al., 2015), and glycated haemoglobin (Roth et al., 2016) are known to change with age, and corpuscular volume has been associated with cognitive functioning (Gamaldo et al., 2013). Lifestyle factors involving alcohol and smoking impact various biological and bodily ages (Karasik et al., 2005), including that of the brain (Cole, 2020; Lange et al., 2020; Wrigglesworth et al., 2021). Elevated blood pressure and other cardiovascular risk factors have established associations with increased brain age (Lange et al., 2020; Lange A. Marie et al., 2021), and an increase in predicted age has been observed in patients with diabetes (Franke et al., 2013). These results need to be assessed carefully due to the large sample size and relatively small effect sizes (Sullivan and Feinn, 2012), but in sum we believe they indicate that our model makes biologically meaningful predictions. Further, transferring the model to unseen datasets comprised of patients with different clinical conditions and matched healthy peers revealed both high accuracy in terms of age prediction and higher deltas among patients with brain disorders, in line with previous studies (Kaufmann et al., 2019). Importantly, the data in the case-control datasets were obtained from scanners not included in the training set, supporting that the model generalizes to previously unseen scanners; a highly valuable asset.

As a further attempt to understand the SFCN-reg model we correlated variance in the prediction-space with variation in imaging-space after standard preprocessing. Explaining the predictions of deep neural networks is inherently hard, and the plethora of methods for interpretation that exist are largely tailored to classification models. Additionally, they often rely on humanly discernable features for validations, such as faces or clearly identifiable objects. Furthermore, in our analyses we were not interested in interpreting the predicted age per se, but rather its deviation from the chronological age of the participant, the brain age delta. Given these factors, we resolved to correlating the brain age delta with standard measures derived from MRI images, such as cortical measures and subcortical volumes, using linear methods. The delta was associated with surface-wide patterns of cortical thinning, as well as volumetric decreases of multiple subcortical areas. Such atrophic trajectories have been linked to the ageing brain (Fjell et al., 2015; Walhovd et al., 2005) both in healthy populations (Fjell et al., 2009) and those with brain disorders (Jacobsen et al., 2014; Pini et al., 2016). While these overall patterns are plausible, and provide further confidence in our model, methodological advances are required to precisely describe its inner workings.

Compared to the all-in-one brain age delta, a single number describing the difference between apparent and chronological age, our results showed increased predictive value for MS, AD and MCI when using the internal representations of the SFCN-reg underlying the brain age predictions. This supports the view that for some applications the constituent components of the singular brain age delta are relevant beyond the age prediction alone (Smith et al., 2020). The innate ability of deep neural networks to form abstractions of the brain at different spatial resolutions throughout the layers of the model may help disentangle individual differences in neurodevelopmental and age-related processes related to complex disorders and traits. Furthermore, we see this result as evidence that deep learning models trained to predict age in large multisite datasets constitute excellent starting points for transfer learning, which can subsequently be fine-tuned to a variety of tasks.

There is an ongoing discussion in the field on whether brain age models that are precise, or those that allow for sufficient variance in their single-subject predictions, are the most useful in a downstream analysis of behavioural and clinical traits (Bashyam et al., 2021; 2020; Hahn et al., 2021). An argument for a model which allows for more variation (a 'looser' fit) is that this would more accurately depict brain age as a complex process which appears differently in different individuals. One challenge with this approach is that the brain age delta is a residual, and recognizing what portion of this error comes from biological variation and what is modelling imprecision is practically unfeasible. As more complex models such as deep CNNs become competitive for brain age modelling, it becomes possible to minimize the overall model error, including the methodological portion, while still allowing the model to accurately represent the necessary biological variability internally.

There are some limitations of the present study which we acknowledge here. Given the computational cost of training complex deep learn-

ing models on such a large dataset we restrained our study to a limited number of possible models, both in terms of model architectures and hyperparameter settings. With further increases in sample sizes and diversity, deeper architectures may be sensible, as the risk of overfitting is directly alleviated by the larger datasets. While we refer to our dataset as diverse, most participants included in the study are of white European background. However, this limitation is not specific to the present study, and global collaborations are needed to build models that generalize across cultural and genetic backgrounds (Thompson et al., 2020). Relatedly, the current investigation included T1-weighted MRI data only. Future integration of information spanning various imaging modalities may increase both age prediction accuracy and the sensitivity and specificity of various biological and clinical traits and conditions (Lange A. Marie et al., 2021; Richard et al., 2018; Rokicki et al., 2021; Smith et al., 2020). To evaluate the clinical relevance of the brain age model in the context of disorders we employed transfer learning in a strict way, by training LASSO models on top of the SFCN architecture. This setup increases the interpretability of the transfer process as the lower-level representations of our model used as predictors capture age-related features of the brain. In turn, this allows for the interpretation that illnesses that are predictable by the model must also rely on these representations, and thus implicitly relate to age. There are multiple steps which could have been taken instead to maximize predictability, a natural starting point is to fine-tune the entire model (Bashyam et al., 2020). Lastly, in order to reach broad adoption of these models and, ultimately, approach clinical usability, a better understanding of the regional patterns driving the prediction, their specific biological significance and how it changes across time and contexts is needed (Vidal-Piñeiro et al., 2021).

In conclusion, we have trained multiple variants of a deep neural network to predict brain age on a large and heterogeneous sample of raw structural MRI data, and observed distinct differences in their ability to generalize to unseen samples and scanners. The predictions of our best model were linked to biochemical biomarkers, cardiovascular risk factors, smoking and alcohol intake, among others. Using transfer learning, we demonstrated that clinical conditions with a neurodevelopmental or neurodegenerative aetiology were predictable by our model, initially trained to predict age. Jointly, these findings add to the growing literature documenting the tremendous potential of advanced techniques for statistical learning to decode biologically and clinically relevant information from brain MRI scans.

Materials and Methods

Data

All data sets used in the present study have been obtained from previously published studies which have been approved by their respective institutional review board or relevant research ethics committee.

Reference dataset

The reference dataset used for training the brain age models was T1-weighted MRI scans derived from 21 non-overlapping and publicly available datasets (total n=53542; female n=27715) of healthy individuals, with ages ranging from 3 to 95 years (Fig. 1 and Supplementary Table 1). The age distribution can be seen in Fig. 1a. The younger age-range (3–30 years) was mainly composed of participants from multiple different datasets. Though the older age-range (40–80) also included multiple datasets, UKBB (obtained from the data repository under accession number 27412) accounted for most of these participants. The most sparsely populated age-ranges were in the very young (147 participants with age ≤ 5), very old (17 participants with age ≥ 85), and in midlife (42 participants with $35 \leq$ age < 45). For each of the datasets, participants that had one or more psychiatric, neurological and/or other relevant diagnoses (Supplementary Table 4), and those withdrawn from the respective study were excluded before model training. In addition, for participants having multiple brain scans, the baseline data were used, such that in the final dataset each data point represents a unique participant drawn from a normative population.

External dataset

To evaluate the generalizability of our trained brain age prediction model, an external dataset was collected (Supplementary Table 2). This dataset included the IXI project (Supplementary Table 5) and healthy controls from the datasets underlying the clinical data described below (total n=2553). Importantly, the external dataset contained images generated by scanners not used in the reference dataset, and subsequently unknown to the models during training and validation. This dataset can be seen as having a more uniform age distribution (Fig. 2a), meaning that our test would capture whether any given model relies too heavily on information observed in the training data.

Clinical data

The clinical data consisted of six patient cohorts diagnosed with MCI, AD, MS, SCZ, a mix of psychotic diagnoses and mood disorders, where the latter was a combination of two cohorts with depression and bipolar disorder (Fig. 2a and Supplementary Table 3). The individual cohorts were compiled from ADNI, AIBL, and multiple scanners at the Oslo University Hospital (Supplementary Tables 5 and 6). In addition to the patients, we used healthy controls from the same scanners in the external dataset to create matched control groups for the clinical predictions.

Quality control

To ensure data quality, we executed a quality control (QC) pipeline, consisting of checking whether any of the image preprocessing steps failed, and a manual control via visual inspection. To take advantage of as much data as possible this manual control was lenient, removing samples where either a significant portion of the brain was missing, or where the orientation of the head was dramatically off, and resulted in dropping only 39/53581 participants in the reference dataset (0.07%) and 1/2554 participants in the external dataset (0.03%).

Image preprocessing

We first performed skull-stripping with the FreeSurfer 5.3 auto-recon pipeline (Sgonne et al., 2004) to produce a brainmask, minimizing the amount of non-brain information in the data, then reoriented the images to the standard FSL (Jenkinson et al., 2012) orientation using fslreorient2std. The resulting images were linearly registered to the MNI152 space using FLIRT (Jenkinson, 2001) with linear interpolation and the default 1 mm FSL template (version 6.0). We cropped away borders of [6:173, 2:214, 0:160] voxels, in the sagittal, coronal and axial dimensions respectively. This cropping yielded the smallest cuboid with marginal loss of brain-related information across the dataset, minimizing the memory footprint of the models during training. As a last preprocessing step the voxel intensity values of all brain images were normalized to the range [0, 1].

Brain age models

The state-of-the-art network architecture, the Simple Fully Convolutional Network (SFCN) (Peng et al., 2021), was implemented as the backbone in all our brain age models. The SFCN architecture consists of a VGG (Simonyan and Zisserman, 2015)-like structure, with five repeated convolutional blocks, each with a three-dimensional convolutional layer with a filter size of (3, 3, 3), a batch normalization layer, rectified linear activation function (ReLU) activation, and a max-pooling layer with a pooling size of (2, 2, 2) (Fig. 1b). The model then has a channel-wise convolutional layer, a last batch normalization layer and a global average pooling layer. From this backbone we defined three end-to-end variants for brain age prediction: The original soft classification model (SFCN-sm), a regression variant with a single output

node with a linear activation (SFCN-reg) and a ranking model from the general age-regression literature (SFCN-rank) (Chen et al., 2017), an approach which has also been successful for brain age predictions (Xia et al., 2021). All the model definitions rely on a matrix X of dimensions $[N, h, w, d]$ containing MRI images as input, an N-dimensional vector age=$[\text{age}_0, \text{age}_1, \dots, \text{age}_{N-1}]$ containing the ground truth ages of the participants to compute its loss, and are ultimately able to produce an N-dimensional vector $\hat{\text{age}} = [\hat{\text{age}}_0, \hat{\text{age}}_1, \dots, \hat{\text{age}}_{N-1}]$ with a single brain age prediction per participant (although this is not necessarily the direct output of the model).

SFCN-sm

The soft classification variant formulates the age regression problem as a multiclass classification problem, by having $m = \lceil \max(\text{age}) \rceil - \lfloor \min(\text{age}) \rfloor$ output neurons, where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceiling and floor operators respectively. It is denoted as soft because it uses a target vector per participant generated by a normal distribution centered around the ground truth age, instead of the one-hot encoding used in regular classification

$$y_i = \mathcal{N}(\text{age}_i, 1).$$

The predictions of the model are similarly a vector of length m with a softmax activation

$$\hat{y}_i = [\hat{y}_{i,0}, \hat{y}_{i,1}, \dots, \hat{y}_{i,(m-1)}], \sum_{j=0}^{m-1} \hat{y}_{i,j} = 1.$$

The loss for a single datapoint is the Kullback-Leibler divergence between the two vectors

$$\text{loss}_i = \text{KL}(y_i || \hat{y}_i).$$

The final age-prediction of a participant is calculated as a weighted sum of the prediction vector

$$\hat{age}_i = \lfloor \min(\text{age}) \rfloor + \sum_{j=0}^{m-1} j \hat{y}_{i,j}.$$

SFCN-reg

The regression variant has a single output neuron predicting a single value $\hat{y}_i$ per participant, limited to the range (min(age), max(age)) with a bounded ReLU activation. The value of $\hat{y}_i$ can be used directly as the predicted age for a participant, $\hat{age}_i = \hat{y}_i$. During training, the model optimizes the mean squared error.

SFCN-rank

The ranking model formulates the age regression problem as a set of binary "*Is participant X older than age y?*"-questions. Like the soft classification model, the model has $m = \lceil \max(\text{age}) \rceil - \lfloor \min(\text{age}) \rfloor$ output neurons, each representing one such binary question. The target vector for a participant is a binary vector of length m, with a 1 in bins corresponding to ages younger than the participants age and 0 in the rest.

$$y_i = [y_{i,0}, y_{i,1}, \dots, y_{i,(m-1)}], y_{i,x} = \begin{cases} 1 & \lfloor \min(\text{age}) \rfloor + x \leq \text{age}_i \\ 0 & else \end{cases}.$$

For each participant the model predicts a vector of length m, where each output neuron is limited to the range $[0, 1]$ by a sigmoid activation

$$\hat{y}_i = [\hat{y}_{i,0}, \hat{y}_{i,1}, \dots, \hat{y}_{i,(m-1)}], \hat{y}_{i,x} \in [0, 1].$$

The model optimizes the mean binary cross entropy across all the output neurons

$$\text{loss}_i = -\frac{1}{m} \sum_{j=0}^{m} y_{i,j} \log(\hat{y}_{i,j}) + (1 - y_{i,j}) log(1 - \hat{y}_{i,j}).$$

To calculate a predicted age for the model we sum up the number of age bins for which the model predicts that a participant is older than the given age (it is worth noting that this limits the model to integer predictions)

$$\hat{age}_i = \lfloor \min(\text{age}) \rfloor + \sum_{j=0}^{m} \mathbb{1}_{x \geq 0.5}(\hat{y}_{i,j}).$$

Brain age model training and comparison

All brain age models were trained on 2 NVIDIA V100 GPUs with 32GB memory, using the Keras Chollet (2015) interface of Tensorflow 2.3 Abadi et al. (0000)) on top of cuda 10.0. Using a batch size of 14 the models took approximately 1 second per step, translating into roughly 45 minutes per epoch or about 2.5 days per full training session. To train the brain age models, 80% (n=42829) of the reference dataset was used for model building (training and validation) and 20% (n=10713) for testing. Among the data for model building 80% (n=34285) and 20% (n=8544) were used for training and validation of the models, respectively (Fig. 1c and Supplementary Table 1). Before these splits, the data was stratified by age and original study to ensure that all subsets had resembling age distributions and came from multiple scanners. Given the great computational cost of model training, determining optimal hyperparameter values by searching over the full configuration space for each model is impractical. Instead, we employed post-hoc heuristics, i.e., tweaking the models based on previous runs. For each training run we trained the model from scratch (with randomly initialized parameters) for 80 epochs, optimized by vanilla stochastic gradient descent, employing an annealing, step-wise, learning rate schedule. This schedule had three steps, reducing the learning rate by a factor of 3 after epochs 20, 40 and 60. The initial learning rate was found independently for each model variant using a learning rate sweep (Smith, 2017) (Supplementary Figure 7). We used mean absolute error (MAE) on the validation split to determine the best epoch for each run. We also report RMSE, R, R^2 for all models in Table 1, and to enable comparisons with other studies with possibly different age ranges the normalized measures normalized RMSE (nRMSE)

$$\text{nRMSE} = \frac{\text{RMSE}}{\max(\text{age}) - \min(\text{age})}$$

and Relative Absolute Error (RAE)

$$\text{RAE} = \frac{\sum_{i=0}^{N} |\hat{y}_i - y_i|}{\sum_{i=0}^{N} |y_i - \bar{y}|}$$

The first model we trained was SFCN-sm with the hyperparameters specified in the original SFCN paper (Peng et al., 2021). Seeing that this model was underfitting we relaxed the regularization for a second run of the same model, and subsequently a third. The two hyperparameters we tuned in this process were the weight decay, and the dropout rate between the two final layers of the model. Having trained three soft classification models, we moved on to train three regression models and three ranking models using these same heuristics (Supplementary Table 8). To select a candidate model for each variant we compared the MAEs on the validation split. In the final model selection, we compared the MAEs of the candidate models for each variant on the test set and the external dataset.

Correlating the brain age delta with imaging measures

To investigate the influence of different brain regions and voxel-wise patterns on our predictions, we performed two post-hoc analyses. First, we carried out a qualitative comparison of groups with unusually low and high brain age deltas, stratified by age. To minimize the impact of scanner effects, both in prediction- and voxel-space, while retaining large enough groups to enable meaningful comparisons, we performed this analysis in the test-portion of UKBB. We also executed the analysis

independently for each sex, to ensure similarities within the groups. For each age bin $b \in [50, 55, 60, 65, 70, 75]$ we selected 10 participants with $age = b \pm 0.5$ with the lowest and highest brain age delta, and compared them with a group of participants the same age with the smallest absolute delta. For each of these groups we created an average brain using FreeSurfer's make_average_brain. The average brains are seen in Supplementary Figure 8.

Additionally, we performed a quantitative analysis correlating variation in brain age delta with structural imaging measures derived by FreeSurfer at the vertex (for cortical features) and atlas level (for subcortical features). This analysis also used subjects from a single scanner to minimize potential biases in the results. Being less reliant on a very large dataset, we were able to use data from the largest scanner in the external dataset, the Oslo GE 750 scanner (n=876). For each subject we computed cortical thickness for each vertex, and subcortical volumes, using FreeSurfer's recon-all pipeline and the default FreeSurfer atlas. For each of these measures we computed the Pearson correlation with the brain age delta across all healthy subjects from the given scanner. The vertex-wise and subcortical correlations are shown in Fig. 3, Supplementary Figure 9 and Supplementary Table 11.

Post-hoc generalization analysis

To study the causes of the differences in generalization, we designed an experiment to isolate the underlying sources of this error. Based on previous knowledge of the problems of new scanners, and the predictions of the models at different ages (Fig. 4) we specifically targeted two possible sources: Differences in population, represented by different distributions of age and sex, and data coming from unknown scanners. For each source we sampled an artificial, bootstrapped dataset based on our existing data. For the "Unknown population" dataset we sampled participants from the test set (originating from the reference dataset), to match the empirical age and sex distribution of the external dataset. Similarly, for the "Unknown scanners" dataset we sampled participants from the external dataset (coming from scanners unknown to the model) to match the age and sex distribution of the test set (and thus also the training set). The idea behind both datasets is to isolate a single source of generalization error. For robustness, we bootstrapped each of these two artificial datasets 100 times and reported the mean MAE achieved by the different models. Each sample was drawn probabilistically, with replacement, with the probability of drawing participant x of age x_a and sex x_s from dataset source based on the age and sex distribution of dataset target given by

$$p(x) = \frac{\hat{P}_{target}(x_a, x_s)}{\hat{P}_{source}(x_a, x_s)}$$

where $\hat{P}_{dataset}(x_a, x_s)$ denotes the proportion of participants in dataset $dataset$ with age x_a and sex x_s. In "Unknown population", the test set plays the role of source and the external dataset is the target, while this is switched for "Unknown scanners".

Phenome-wide association study

In the PheWAS we calculated the univariate correlation between the brain age delta and 402 phenotypic variables from the UKBB, manually divided into thirteen thematically defined categories for interpretability (Supplementary Table 7). We performed this analysis in the UKBB portion of the test split (n=8066), and used all the variables available to us at the time. We encoded all phenotypic variables according to the PHESANT (Millard et al., 2018) datatypes, and removed non-informative levels (Supplementary Table 12) based on the UKBB coding schemes. Additionally, we re-coded the ordinal variables as categorical or continuous by hand (Supplementary Table 13). Variables which were impossible to model (i.e. singular or all missing values) were discarded. We then fitted a linear model per variable, modelling the delta as a function of the given covariate, age and sex, using the Python statsmodels API (Seabold and Perktold, 2010). All continuous variables were standardized using a z-score normalization pre-modelling, such that the reported effect sizes refer to the change in brain age delta associated with a one standard deviation increase in the given variable. For assessing the significance of the associations, we computed a Bonferroni-corrected p-value threshold $p_{bonferroni} = 0.05/394 = 1.12 \times 10^{-4}$.

Transfer learning to clinical samples

In the transfer learning analysis we trained multiple binary models to predict whether a participant had a given diagnosis or belonged to the control group, based on various levels of information from the brain age model. For this purpose, we used the clinical dataset which was previously unseen by the model, and matched controls groups drawn from the external dataset used in the generalization test.

Control matching

To avoid biases in the case-control datasets we drew the subsets of controls independently for each scanner in the patient dataset, matching empirical distributions of age and sex in the corresponding case subset. For each disorder, for each scanner, the control group was created by drawing n_{case} controls, without replacement, using the sampling procedure described for the post-hoc generalization analysis.

Feature extraction

To generate the feature vectors for each participant we used the trained SFCN-reg model as an encoder, up until and including the global average-pooling layer. Running a single MRI through the model up until this point results in a 64-dimensional vector representing the original image. Each of the dimensions $n = [0, \dots, 63]$ in this space represent a high level feature of the brain, and each participant $X_i = [X_{i,0}, \dots, X_{i,63}]$ is encoded as a point in this space. When used as a predictor in the subsequent modelling phase, each of these 64 dimensions were treated as an independent variable.

Modelling

For each disorder we compared three different LASSO models, all trained and evaluated using the following general procedure, but on different covariate sets. We first stratified the given dataset on disorder, age and sex, respectively, and split it into 5 folds. We performed an outer cross-validation over these splits to allow us to have an out-of-sample prediction for each participant. When training a model on the training folds we performed an inner cross-validation to find the optimal value of the penalty parameter λ. The nested cross-validation procedure is illustrated in Supplementary Figure 5. Having found λ, we retrained the model on all the data from the four training folds. The models were implemented using sklearn's LogisticRegression Pedregosa et al. (0000)) with an l_1-penalty. Having the out-of-sample predictions for all the participants allowed us to calculate and compare AUCs based on the entire case-control dataset for the given disorder.

In addition to training the LASSO-model based on the brain age features, we trained an MLP using Keras with the same inputs. We did not optimize hyperparameters for this model, but observed similar results as the best LASSO models with the initial configuration (Supplementary Figure 6). The main benefit of the MLP is that it does not require a two-step process for the clinical prediction models, first processing the images with the encoder and then doing a prediction via a separate API, but can be implemented as an end-to-end binary classifier in Keras taking MRIs as inputs, and thus are more accessible for use by others.

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 independent datasets, and their origins, is provided in Supplementary Table 5.

Code availability

All of the trained brain age models and a pipeline for preprocessing images is released in our GitHub repo at http://www.github.com/estenhl/pyment-public

Funding sources

This work was funded by the UiO:LifeScience Convergence Environment (project: 4MENT), The Research Council of Norway (302854, 223273, 249795, 298646, 300767), the South-Eastern Norway Regional Health Authority (2014097, 2016083, 2018037, 2018076, 2019101), Stiftelsen Kristian Gerhard Jebsen, ERA-Net Cofund through the ERA PerMed project 'IMPLEMENT', and the European Research Council under the European Union's Horizon 2020 research and Innovation program (ERC StG, Grant 802998) and the Wellcome Trust grant (215698/Z/19/Z). T.W. gratefully acknowledges support from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 895011. A.F.M. gratefully acknowledges support from the European Research Council (ERC, grant number 10100118) and the Dutch Organisation for Scientific Research (016.156.415). A.M.d.L. was supported by the Swiss National Science Foundation (PZ00P3_193658). T.K. was supported by the Research Council of Norway (276082, 323961). Part of the computation was performed on the Norwegian high-performance computation resources, sigma2 (NN9767K/NS9769K). Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL) databases (adni.loni.usc.edu), and the Pediatric Imaging, Neurocognition and Genetics (PING) study database (chd.ucsd.edu/research/ping-study.html, now shared through the NIMH Data Archive (NDA)). The investigators within the ADNI, AIBL, and PING studies contributed to the design and implementation of ADNI, AIBL, 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.

Declaration of Competing Interest

Declarations of interest: None

Credit authorship contribution statement

Esten H. Leonardsen: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Han Peng:** Conceptualization, Methodology, Software, Validation, Visualization, Writing – review & editing. **Tobias Kaufmann:** Conceptualization, Methodology, Software, Validation, Visualization, Writing – review & editing. **Ingrid Agartz:** Data curation, Validation, Writing – review & editing. **Ole A. Andreassen:** Data curation, Funding acquisition, Validation, Writing – review & editing. **Elisabeth Gulowsen Celius:** Data curation, Validation, Writing – review & editing. **Thomas Espeseth:** Data curation, Validation, Writing – review & editing. **Hanne F. Harbo:** Data curation, Validation, Writing – review & editing. **Einar A. Høgestøl:** Data curation, Investigation, Validation, Writing – review & editing. **Ann-Marie de Lange:** Data curation, Validation, Writing – review & editing. **Andre F. Marquand:** Investigation, Methodology, Validation, Writing – review & editing. **Didac Vidal-Piñeiro:** Investigation, Methodology, Validation, Writing – review & editing. **James M. Roe:** Investigation, Methodology, Validation, Writing – review & editing. **Geir Selbæk:** Data curation, Validation, Writing – review & editing. **Øystein Sørensen:** Investigation, Methodology, Validation, Writing – review & editing. **Stephen M. Smith:** Investigation, Methodology, Software, Validation, Writing – review & editing. **Lars T. Westlye:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft. **Thomas Wolfers:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft. **Yunpeng Wang:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119210.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 19

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., Calhoun, V., 2021. Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. Nat Commun 12 (1), 353.

Alfaro-Almagro, F., McCarthy, P., Afyouni, S., Andersson, J.L.R., Bastiani, M., Miller, K.L., Nichols, T.E., Smith, S.M., 2021. Confound modelling in UK Biobank brain imaging. Neuroimage 224, 117002.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. Neuroimage 145, 137–165.

Ashpole, N.M., Sanders, J.E., Hodges, E.L., Yan, H., Sonntag, W.E., 2015. Growth hormone, insulin-like growth factor-1 and the aging brain. Exp. Gerontol. 68, 76–81.

Bashyam, V., Shou, H., Davatzikos, C., 2021. Reply: from 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. Brain 144 (3), 32.

Bashyam, V.M., Erus, G., Doshi, J., Habes, M., Nasrallah, I.M., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., Launer, L.J., Masters, C.L., Maruff, P., Zhuo, C., Vlzke, H., Johnson, S.C., Fripp, J., Koutsouleris, N., Satterthwaite, T.D., Wolf, D., Gur, R.E., Gur, R.C., Morris, J., Albert, M.S., Grabe, H.J., Resnick, S., Nick, B.R., Wolk, D.A., Shou, H., Davatzikos, D., 2020. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. Brain 143 (7), 2312–2324.

Bengio, Y., 2012. Deep Learning of Representations for Unsupervised and Transfer Learning. In: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, pp. 17–36.

Boyle, R., Jollans, L., Rueda-Delgado, L.M., Rizzo, R., Yener, G.P., McMorrow, J.P., Knight, S.P., Carey, D., Robertson, I.H., Emek-Sava, D.D., Stern, Y., Kenny, R.A., Whelan, R., 2021. Brain-predicted age difference score is related to specific cognitive functions: a multi-site replication analysis. Brain Imaging Behav 15 (1), 327–345.

Butler, E.R., Chen, A., Ramadan, R., Le, T.T., Ruparel, K., Moore, T.M., Satterthwaite, T.D., Zhang, F., Shou, H., Gur, R.C., Nichols, T.E., Shinohara, R.T., 2021. Pitfalls in brain age analyses. Hum Brain Mapp 42 (13), 4092–4101.

Chen, S., Zhang, C., Dong, M., Le, J., Rao, M., 2017. Using ranking-CNN for age estimation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 742–751.

Chollet, F., 2015. Keras.

Cole, J.H., 2020. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. Neurobiol. Aging 92, 34–42.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends Neurosci. 40 (12), 681–690.

Cole, J.H., Poudel, R.P.K., Tsagkrasoulis, D., Caan, M.W.A., Steves, C., Spector, T.D., Montana, G., 2017. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage 163, 115–124.

Dinsdale, N.K., Bluemke, E., Smith, S.M., Arya, Z., Vidaurre, D., Jenkinson, M., Namburete, A.I.L., 2021. Learning patterns of the ageing brain in MRI using deep convolutional networks. Neuroimage 224, 117401.

Dinsdale, N.K., Jenkinson, M., Namburete, A.I.L., Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Kevin, Z.S., Racoceanu, D., Joskowicz, L., 2020. Unlearning Scanner Bias for MRI Harmonisation. In: Medical Image Computing and Computer Assisted Intervention 'MICCAI' 2020, pp. 369–378.

Dockès, J.m., Varoquaux, G., Poline, J.B., 2021. Preventing dataset shift from breaking machine-learning biomarkers. Gigascience 10 (9).

Elliott, M.L., Belsky, D.W., Knodt, A.R., Ireland, D., Melzer, T.R., Poulton, R., Ramrakha, S., Caspi, A., Moffitt, T.E., Hariri, A.R., 2019. Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. Mol Psychiatry 1–10.

Fjell, A.M., Grydeland, H., Krogsrud, S.K., Amlien, I., Rohani, D.A., Ferschmann, L., Storsve, A.B., Tamnes, C.K., Sala-Llonch, R., Due-Tønnessen, P., Bjørnerud, A., Sølsnes, A.E., Håberg, A.K., Skranes, J., Bartsch, H., Chen, C.H., Thompson, W.K., Panizzon, M.S., Kremen, W.S., Dale, A.M., Walhovd, K.B., 2015. Development and aging of cortical thickness correspond to genetic organization patterns. Proc. Natl. Acad. Sci. U.S.A. 112 (50), 15462–15467.

Fjell, A.M., Walhovd, K.B., 2010. Structural brain changes in aging: courses, causes and cognitive consequences. Rev Neurosci 21 (3), 187–222.
Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M., 2009. One-year brain atrophy evident in healthy aging. J. Neurosci. 29 (48), 15223–15231.
Franke, K., Gaser, C., 2019. Ten years of brainAGE as a neuroimaging biomarker of brain aging: what insights have we gained? Front. Neurol. 10.
Franke, K., Gaser, C., Manor, B., Novak, V., 2013. Advanced brainAGE in older adults with type 2 diabetes mellitus. Front. Aging Neurosci. 5.
Franke, K., Luders, E., May, A., Wilke, M., Gaser, C., 2012. Brain maturation: predicting individual brainAGE in children and adolescents using structural MRI. Neuroimage 63 (3), 1305–1312.
Franke, K., Ziegler, G., Klppel, S., Gaser, C., 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. Neuroimage 50 (3), 883–892.
Gamaldo, A.A., Ferrucci, L., Rifkind, J., Longo, D.L., Zonderman, A.B., 2013. Relationship between mean corpuscular volume and cognitive performance in older adults. J Am Geriatr Soc 61, 84–89.
Gaser, C., Franke, K., Klåppel, S., Koutsouleris, N., Sauer, H., Initiative, A.s., 2013. BrainAGE in mild cognitive impaired patients: predicting the conversion to alzheimers disease. PLoS ONE 8 (6), e67346.
Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M., Peng, H., 2021. Optimising a simple fully convolutional network for accurate brain age prediction in the PAC 2019 challenge. Front Psychiatry 12, 658.
Hahn, T., Fisch, L., Ernsting, J., Winter, N.R., Leenings, R., Sarink, K., Emden, D., Kircher, T., Berger, K., Dannlowski, U., 2021. From 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. Brain 144 (3), 31.
Han, L.K.M., Dinga, R., Hahn, T., Ching, C.R.K., Eyler, L.T., Aftanas, L., Aghajani, M., Aleman, A., Baune, B.T., Berger, K., Brak, I., Filho, G.B., Carballedo, A., Connolly, C.G., Couvy-Duchesne, B., Cullen, K.R., Dannlowski, U., Davey, C.G., Dima, D., Duran, F.L.S., Enneking, V., Filimonova, E., Frenzel, S., Frodl, T., Fu, C.H.Y., Godlewska, B.R., Gotlib, I.H., Grabe, H.J., Groenewold, N.A., Grotegerd, D., Gruber, O., Hall, G.B., Harrison, B.J., Hatton, S.N., Hermesdorf, M., Hickie, I.B., Ho, T.C., Hosten, N., Jansen, A., Khler, C., Kircher, T., Klimes-Dougan, B., Krmer, B., Krug, A., Lagopoulos, J., Leenings, R., MacMaster, F.P., MacQueen, G., McIntosh, A., McLellan, Q., McMahon, K.L., Medland, S.E., Mueller, B.A., Mwangi, B., Osipov, E., Portella, M.J., Pozzi, E., Reneman, L., Repple, J., Rosa, P.G.P., Sacchet, M.D., Smann, P.G., Schnell, K., Schrantee, A., Simulionyte, E., Soares, J.C., Sommer, J., Stein, D.J., Steinstrter, O., Strike, L.T., Thomopoulos, S.I., Tol, M.J., Veer, I.M., Vermeiren, R.R.J.M., Walter, H., Wee, N.J.A., Werff, S.J.A., Whalley, H., Winter, N.R., Wittfeld, K., Wright, M.J., Wu, M.J., Vlzke, H., Yang, T.T., Zannias, V., Zubicaray, G.I., Zunta-Soares, G.B., Ab, C., Alda, M., Andreassen, O.A., Bnm, E., Bonnin, C.M., Canales-Rodriguez, E.J., Cannon, D., Caseras, X., Chaim-Avancini, T.M., Elvsshagen, T.r., Favre, P., Foley, S.F., Fullerton, J.M., Goikolea, J.M., Haarman, B.C.M., Hajek, T., Henry, C., Houenou, J., Howells, F.M., Ingvar, M., Kuplicki, R., Lafer, B., Landn, M., Machado-Vieira, R., Malt, U.F., McDonald, C., Mitchell, P.B., Nabulsi, L., Otaduy, M.C., Overs, B.J., Polosan, M., Pomarol-Clotet, E., Radua, J., Rive, M.M., Roberts, G., Ruhe, H.G., Salvador, R., Sarr, S., Satterthwaite, T.D., Savitz, J., Schene, A.H., Schofield, P.R., Serpa, M.H., Sim, K., Soeiro-de Souza, M.G., Sutherland, A.N., Temmingh, H.S., Timmons, G.M., Uhlmann, A., Vieta, E., Wolf, D.H., Zanetti, M.V., Jahanshad, N., Thompson, P.M., Veltman, D.J., Penninx, B.W.J.H., Marquand, A.F., Cole, J.H., Schmaal, L., 2020. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. In: Mol Psychiatry, pp. 1–16. Publisher: Nature Publishing Group.
Høgestøl, E.A., Kaufmann, T., Nygaard, G.O., Beyer, M.K., Sowa, P., Nordvik, J.E., Kolskår, K., Richard, G., Andreassen, O.A., Harbo, H.F., Westlye, L.T., 2019. Cross-Sectional and Longitudinal MRI Brain Scans Reveal Accelerated Brain Aging in Multiple Sclerosis. Front. Neurol., Vol. 0. Publisher: Frontiers
He, S., Pereira, D., DavidPerez, J., Gollub, R.L., Murphy, S.N., Prabhu, S., Pienaar, R., Robertson, R.L., EllenGrant, P., Ou, Y., 2021. Multi-channel attention-fusion neural network for brain age estimation: accuracy, generality, and interpretation with 16,705 healthy MRIs across lifespan. Med Image Anal 72, 102091.
Hofmann, S.M., Beyer, F., Lapuschkin, S., Loeffler, M., Müller, K.R., Villringer, A., Samek, W., Veronica, W.A., 2021. Towards the Interpretability of Deep Learning Models for Human Neuroimaging. bioRxiv, 2021.06.25.449906. Publisher: Cold Spring Harbor Laboratory Section: New Results
Jacobsen, C., Hagemeier, J., Myhr, K.M., Nyland, H., Lode, K., Bergsland, N., Ramasamy, D.P., Dalaker, T.O., Larsen, J.P., Farbu, E., Zivadinov, R., 2014. Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. Journal of Neurology, Neurosurgery & Psychiatry 85 (10), 1109–1115. Publisher: BMJ Publishing Group Ltd Section: Multiple sclerosis
Jenkinson, M., 2001. A global optimisation method for robust affine registration of brain images. Med Image Anal 5 (2), 143–156.
Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62 (2), 782–790.
Johnson, M.H., 2001. Functional brain development in humans. Nat Rev Neurosci 2 (7), 475–483. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Reviews Publisher: Nature Publishing Group
Jonsson, B.A., Bjornsdottir, G., Thorgeirsson, T.E., Ellingsen, L.M., Bragi, W.G., Gudbjartsson, D.F., Stefansson, H.H., Stefansson, K.K., Ulfarsson, M.O., 2019. Brain age prediction using deep learning uncovers associated sequence variants. Nat Commun 10 (1), 5409.
Karasik, D., Demissie, S., Adrienne, C.L., Kiel, D.P., 2005. Disentangling the genetic determinants of human aging: biological age as an alternative to the use of survival measures. The Journals of Gerontology: Series A 60 (5), 574–587.
Kaufmann, T., Meer, D., Doan, N.T., Schwarz, E., Lund, M.J., Agartz, I., Aln, D., Barch, D.M., Baur-Streubel, R., Bertolino, A., Bettella, F., Beyer, M.K., Bøem, E., Borgwardt, S., Brandt, C.L., Buitelaar, J., Celius, E.G., Cervenka, S., Conzelmann, A., Crdova-Palomera, A., Dale, A.M., Quervain, D.J.F., DiCarlo, P., Djurovic, S., Dnm, E.S., Eisenacher, S., Elvsshagen, T.r., Espeseth, T., Fatouros-Bergman, H., Flyckt, L., Franke, B., Frei, O., Haatveit, B., Hberg, A.K., Harbo, H.F., Hartman, C.A., Heslenfeld, D., Hoekstra, P.J., A, H.E., Jernigan, T.L., Jonassen, R., Jnsson, E.G., Kirsch, P., Knm, I., Kolskår, K.K., Landr, N.I., Hellard, S.L., Lesch, K.P., Lovestone, S., Lundervold, A., Lundervold, A.J., Maglanoc, L.A., Malt, U.F., Mecocci, P., Melle, I., Meyer-Lindenberg, A., Moberget, T., Norbom, L.B., Nordvik, J.E., Nyberg, L., Oosterlaan, J., Papalino, M., Papassotiropoulos, A., Pauli, P., Pergola, G., Persson, K., Richard, G.v., Rokicki, J., Sanders, A.M., Selb, G., Shadrin, A.A., Smeland, O.B., Soininen, H., Sowa, P., Steen, V.M., Tsolaki, M., Ulrichsen, K.M., Vellas, B., Wang, L., Westman, E., Ziegler, G.C., Zink, M., Andreassen, O.A., Westlye, L.T., 2019. Common brain disorders are associated with heritable patterns of apparent aging of the brain. Nat Neurosci 22 (10), 1617–1623.
de Lange, A.-M.G., Anatürk, M., Suri, S., Kaufmann, T., Cole, J.H., Griffanti, L., Zsoldos, E., Jensen, D.E.A., Filippini, N., Singh-Manoux, A., Kivimäki, M., Westlye, L.T., Ebmeier, K.P., 2020. Multimodal brain-age prediction and cardiovascular risk: the Whitehall II MRI sub-study. Neuroimage 222, 117292.
de Lang, A.-M.G., Cole, J.H., 2020. Commentary: correction procedures in brain-age prediction. Neuroimage Clin 26, 102229.
Beck, D., de Lang, A.-M.G., Pedersen, M.L., Alnaes, D., Maximov, I.I., Voldsbekk, I., Richard, G., Sanders, A.M., Ulrichsen, K.M., Dørum, E.S., Kolskår, K.K., Høgestøl, E.A., Steen, N.E., Djurovic, S., Andreassen, O.A., Nordvik, J.E., Kaufmann, T., Westlye, L.T., 2021. Cardiometabolic Risk Factors Associated with Brain Age and Accelerate Brain Ageing. Hum Brain Mapp.
Liem, F., Varoquaux, G., Kynast, J., Beyer, F., KharabianMasouleh, S., Huntenburg, J.M., Lampe, L., Rahim, M., Abraham, A., Cameron, C.R., Riedel-Heller, S., Luck, T., Loeffler, M., Schroeter, M.L., Witte, A.V., Villringer, A., Margulies, D.S., 2017. Predicting brain-age from multimodal imaging data captures cognitive impairment. Neuroimage 148, 179–188.
Lu, B., Li, H. X., Chang, Z. K., Li, L., Chen, N. X., Zhu, Z. C., Zhou, H. X., Li, X. Y., Wang, Y. W., Cui, S. X., Deng, Z. Y., Fan, Z., Yang, H., Chen, X., Thompson, P. M., Castellanos, F. X., Yan, C. G., for the Alzheimer's Disease Neuroimaging Initiative, 2020. A Practical Alzheimer Disease Classifier via Brain Imaging-Based Deep Learning on 85,721 Samples.
Millard, M., Davies, N.M., Gaunt, T.R., DaveySmith, G., Tilling, K., 2018. Software application profile: PHESANT: a tool for performing automated phenome scans in UK biobank. Int J Epidemiol 47 (1), 29–35.
Nenadi, I., Dietzek, M., Langbein, K., Sauer, H., Gaser, C., 2017. BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. Psychiatry Research: Neuroimaging 266, 86–89.
Oquendo, M.A., Baca-Garcia, E., Artès-Rodrìguez, A., Perez-Cruz, F., Galfalvy, H.C., Blasco-Fontecilla, H., Madigan, D., Duan, N., 2012. Machine learning and data mining: strategies for hypothesis generation. Mol Psychiatry 17 (10), 956–959.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Scikit-learn: Machine Learning in Python. In: MACHINE LEARNING IN PYTHON, p. 6.
Peng, H., Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M., 2021. Accurate brain age prediction with lightweight deep neural networks. Med Image Anal 68, 101871.
Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., Frisoni, G.B., 2016. Brain atrophy in Alzheimer's disease and aging. Ageing Res. Rev. 30, 25–48.
Ren, Y., Luo, Q., Gong, W., Lu, W., 2019. Transfer learning models on brain age prediction. In: Proceedings of the Third International Symposium on Image Computing and Digital Medicine - ISICDM 2019, pp. 278–282.
Richard, G.v., Kolskår, K., Sanders, A.M., Kaufmann, T., Petersen, A., Doan, N.T., Snchez, J.M., Aln, D., Ulrichsen, K.M., Dnm, E.S., Andreassen, O.A., Nordvik, J.E., Westlye, L.T., 2018. Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. PeerJ 6, e5908.
Rokicki, J., Wolfers, T., Nordh, W., Tesli, N., Quintana, D.S., Aln, D., Richard, G., Lange, A.-M.d., Lund, M.J., Norbom, L., Agartz, I., Melle, I., Nnm, T., Selb, G., Persson, K., Nordvik, J.E., Schwarz, E., Andreassen, O.A., Kaufmann, T., Westlye, L.T., 2021. Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. Hum Brain Mapp 42 (6), 1714–1726.
Roth, J., Müller, N., Lehmann, T., Heinemann, L., Wolf, G., Müller, U.A., 2016. HbA1c and age in non-diabetic subjects: an ignored association? Exp Clin Endocrinol Diabetes 124 (10), 637–642.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. Imagenet large scale visual recognition challenge. Int J Comput Vis 115 (3), 211–252.
Schnack, H.G., Haren, N.E.M., Nieuwenhuis, M., HulshoffPol, H.E., Cahn, W., Kahn, R.S., 2016. Accelerated brain aging in schizophrenia: alongitudinal pattern recognition study. AJP 173 (6), 607–616.
Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and Statistical Modeling with Python. 92–96
Sgonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22 (3), 1060–1075.
SharifRazavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. 806–813
Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv: 1409.1556.

Smith, L. N., 2017. Cyclical Learning Rates for Training Neural Networks. ArXiv: 1506.01186.

Smith, S.M., Elliott, L.T., Alfaro-Almagro, F., McCarthy, P., Nichols, T.E., Douaud, G., Miller, K.L., Peelle, J.E., Lange, F.P., Madan, C., Nyberg, L., 2020. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. Elife 9, e52677.

Smith, S.M., Nichols, T.E., 2018. Statistical challenges in 'big data' human neuroimaging. Neuron 97 (2), 263–268.

Smith, S.M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T.E., Miller, K.L., 2019. Estimation of brain age delta from brain imaging. Neuroimage 200, 528–539.

Stephan, B.C.M., Brayne, C., 2009. Assessing the risk of dementia in the aging population. Nat Rev Neurol 5 (8), 417–418.

Sullivan, G.M., Feinn, R., 2012. Using effect size or why the p value is not enough. J Grad Med Educ 4 (3), 279–282.

Thapar, A., Riglin, L., 2020. The importance of a developmental perspective in psychiatry: what do recent genetic-epidemiological findings show? Mol Psychiatry 25 (8), 1631–1639.

Thompson, P.M., Jahanshad, N., Ching, C.R.K., Salminen, L.E., Thomopoulos, S.I., Bright, J., Baune, B.T., Bertoln, S., Bralten, J., Bruin, W.B., Blow, R., Chen, J., Chye, Y., Dannlowski, U., Kovel, C.G.F., Donohoe, G., Eyler, L.T., Faraone, S.V., Favre, P., Filippi, C.A., Frodl, T., Garijo, D., Gil, Y., Grabe, H.J., Grasby, K.L., Hajek, T., Han, L.K.M., Hatton, S.N., Hilbert, K., Ho, T.C., Holleran, L., Homuth, G., Hosten, N., Houenou, J., Ivanov, I., Jia, T., Kelly, S., Klein, M., Kwon, J.S., Laansma, M.A., Leerssen, J., Lueken, U., Nunes, A., Neill, J.O., Opel, N., Piras, F., Piras, F., Postema, M.C., Pozzi, E., Shatokhina, N., Soriano-Mas, C., Spalletta, G., Sun, D., Teumer, A., Tilot, A.K., Tozzi, L., Merwe, C., VanSomeren, E.J.W., Wingen, G.A., Vlzke, H., Walton, E., Wang, L., Winkler, A.M., Wittfeld, K., Wright, M.J., Yun, J.Y., Zhang, G., Zhang-James, Y., Adhikari, B.M., Agartz, I., Aghajani, M., Aleman, A., Althoff, R.R., Altmann, A., Andreassen, O.A., Baron, D.A., Bartnik-Olson, B.L., MarieBas-Hoogendam, J., Baskin-Sommers, A.R., Bearden, C.E., Berner, L.A., Boedhoe, P.S.W., Brouwer, R.M., Buitelaar, J.K., Caeyenberghs, K., Cecil, C.A.M., Cohen, R.A., Cole, J.H., Conrod, P.J., DeBrito, S.A., Zwarte, S.M.C., Dennis, E.L., Desrivieres, S., Dima, D., Ehrlich, S., Esopenko, C., Fairchild, G., Fisher, S.E., Fouche, J.P., Francks, C., Frangou, S., Franke, B., Garavan, H.P., Glahn, D.C., Groenewold, N.A., Gurholt, T.P., Gutman, B.A., Hahn, T., Harding, I.H., Hernaus, D., Hibar, D.P., Hillary, F.G., Hoogman, M., HulshoffPol, H.E., Jalbrzikowski, M., Karkashadze, G.A., Klapwijk, E.T., Knickmeyer, R.C., Kochunov, P., Koerte, I.K., Kong, X.Z., Liew, S.L., Lin, A.P., Logue, M.W., Luders, E., Macciardi, F., Mackey, S., Mayer, A.R., McDonald, C.R., McMahon, A.B., Medland, S.E., Modinos, G., Morey, R.A., Mueller, S.C., Mukherjee, P., Namazova-Baranova, L., Nir, T.M., Olsen, A., Paschou, P., Pine, D.S., Pizzagalli, F., Rentera, M.E., Rohrer, J.D., Smann, P.G., Schmaal, L., Schumann, G., Shiroishi, M.S., Sisodiya, S.M., Smit, D.J.A., Snm, I.E., Stein, D.J., Stein, J.L., Tahmasian, M., Tate, D.F., Turner, J.A., Heuvel, O.A., Wee, N.J.A., Werf, Y.D., Erp, T.G.M., Haren, N.E.M., Rooij, D., Velzen, L.S., Veer, I.M., Veltman, D.J., Villalon-Reina, J.E., Walter, H., Whelan, C.D., Wilde, E.A., Zarei, M., Zelman, V., 2020. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. Transl Psychiatry 10 (1), 1–28.

Valverde, J.M., Imani, V., Abdollahzadeh, A., DeFeo, R., Prakash, M., Ciszek, R., Tohka, J., 2021. Transfer learning in magnetic resonance brain imaging: a systematic review. Journal of Imaging 7 (4), 66.

Vidal-Piñeiro, D., Wang, Y., Krogsrud, S., Amlien, I., Baarè, W., Bartrès-Faz, D., Bertram, L., Brandmaier, A.M., Drevon, C., Düzel, S., Ebmeier, K., Henson, R.N., Junque, C., Kievit, R.A., Khn, S., Leonardsen, E., Lindenberger, U., Madsen, K., Magnussen, F., Mowinckel, A., Nyberg, L., Roe, J., Segura, B., Smith, S., Suri, S., Westerhausen, R., Zalesky, A., Zsoldos, E., Ageing, T., Walhovd, K., Fjell, A., 2021. Individual Variations in 'Brain Age' Relate to Early Life Factors More than to Longitudinal Brain Change. bioRxiv, 2021.02.08.428915.

Walhovd, K.B., Fjell, A.M., Reinvang, I., Lundervold, A., Dale, A.M., Eilertsen, D.E., Quinn, B.T., Salat, D., Makris, N., Fischl, B., 2005. Effects of age on volumes of cortex, white matter and subcortical structures. Neurobiol. Aging 26 (9), 1261–1270.

Wang, J., Knol, M.J., Tiulpin, A., Dubost, F., Bruijne, M.d., Vernooij, M.W., Adams, H.H.H., Arfan, I.M., Niessen, W.J., Roshchupkin, G.V., 2019. Gray matter age prediction as a biomarker for risk of dementia. PNAS 116 (42), 21213–21218.

Wrigglesworth, J., Ward, P., Harding, I.H., Nilaweera, D., Wu, Z., Woods, R.L., Ryan, J., 2021. Factors associated with brain ageing - a systematic review. BMC Neurol 21 (1), 312.

Xia, T., Chartsias, A., Wang, C., Tsaftaris, S.A., 2021. Learning to synthesise the ageing brain without longitudinal data. Med Image Anal 73, 102169.

Supplementary information

Deep neural networks learn general and clinically relevant representations of the ageing brain

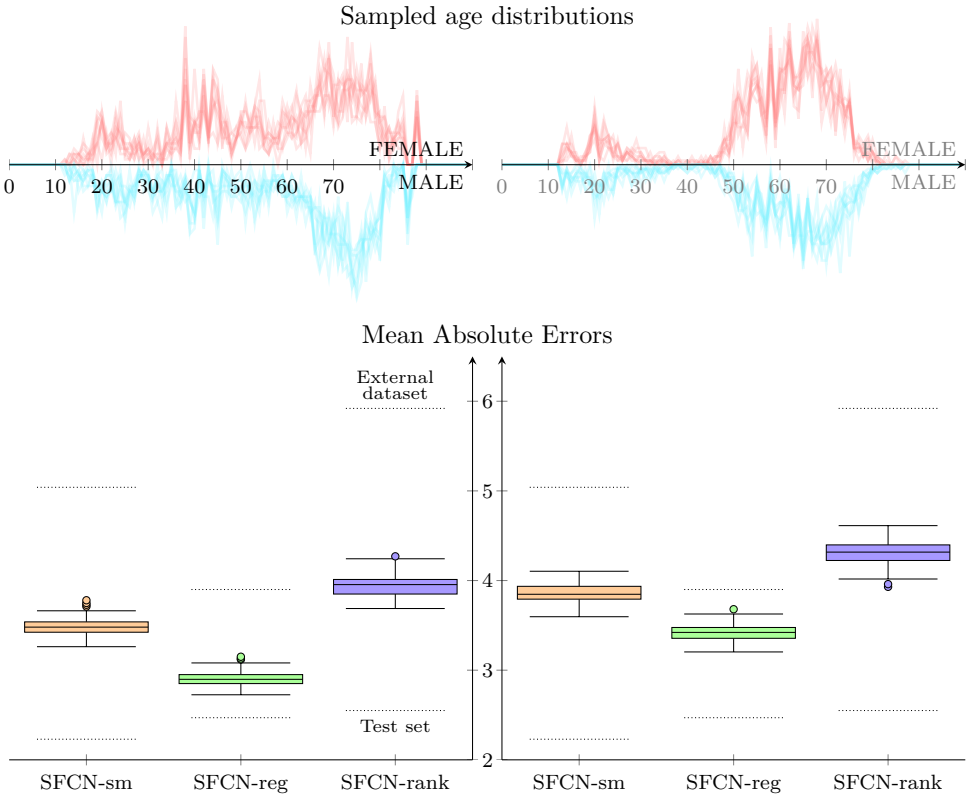
Leonardsen et al.

List of Supplementary Figures

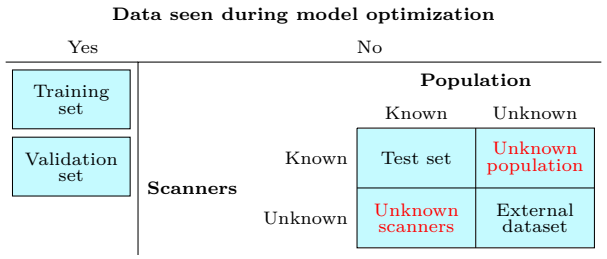
1 Results of the model comparison on the resampled datasets representing a breakdown of the sources of generalization error. 2
2 Predictions of SFCN-reg on the external dataset, separated into the different scanning sites 3
3 Predictions of SFCN-reg on the control sets used in the clinical classifiers 4
4 Brain age delta distributions for the individual patient cohorts versus their individually matched control groups 5
5 The nested cross-validation scheme used for the clinical predictors 6
6 A comparison of the AUCs of the LASSO and CNN clinical predictors . . 7
7 Results of the learning rate sweep for the three model variants 8
8 Average brains for different chronological and brain age groups. 9
9 Correlation between brain age delta and cortical measures 10

List of Supplementary Tables

1 The reference dataset 11
2 The external dataset used for generalization testing 12
3 Patient cohorts used in the clinical classifiers 13
4 Exclusion criteria for the participants in the reference dataset 14
5 Sources of all the datasets used in the study 17
6 Scanning protocols used in the datasets 18
7 Columns used in the PheWAS 21
8 The hyperparameter settings tested for each of the three model variants . 22
9 Results of the final model comparison 23
10 Model performance in the two test datasets 24
11 Correlation between subcortical volumes and brain age delta 26

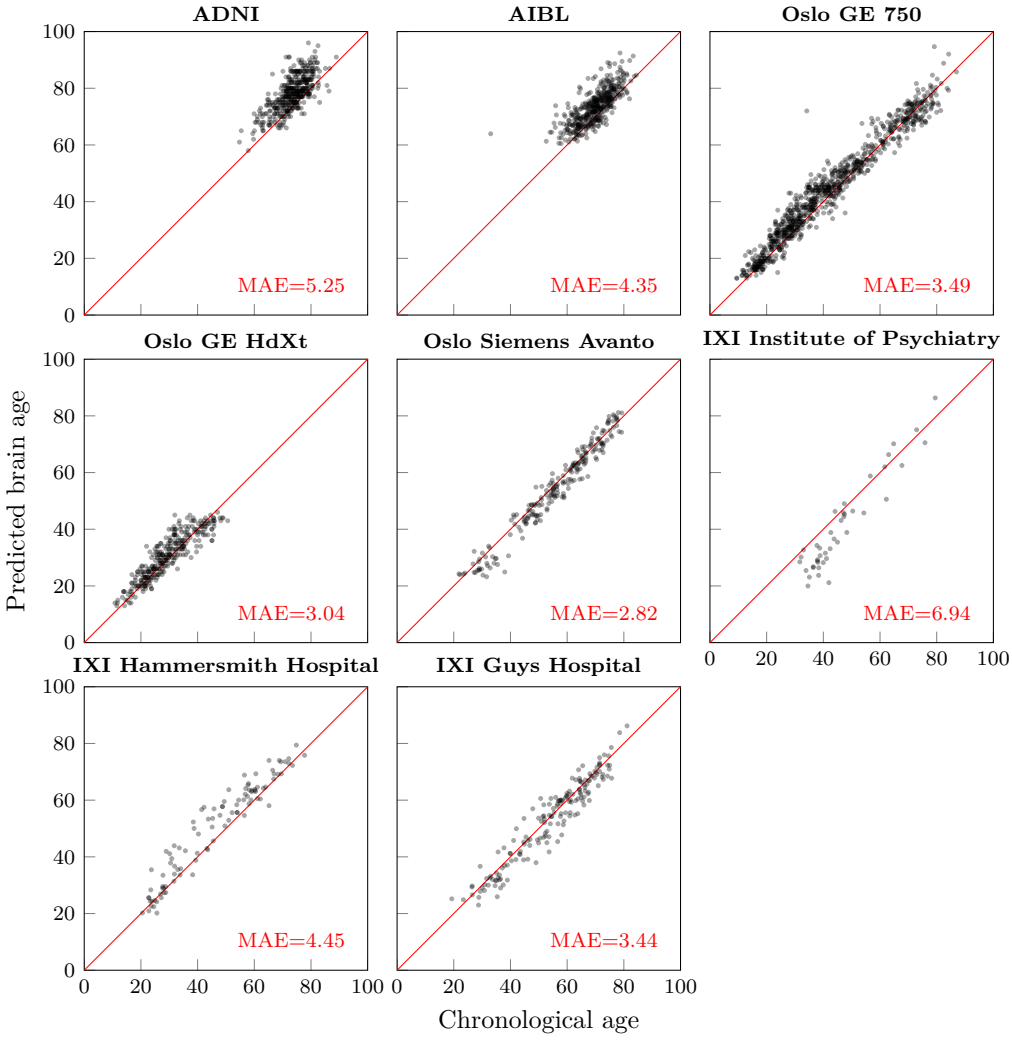


(a) Unknown population **(b) Unknown scanners**

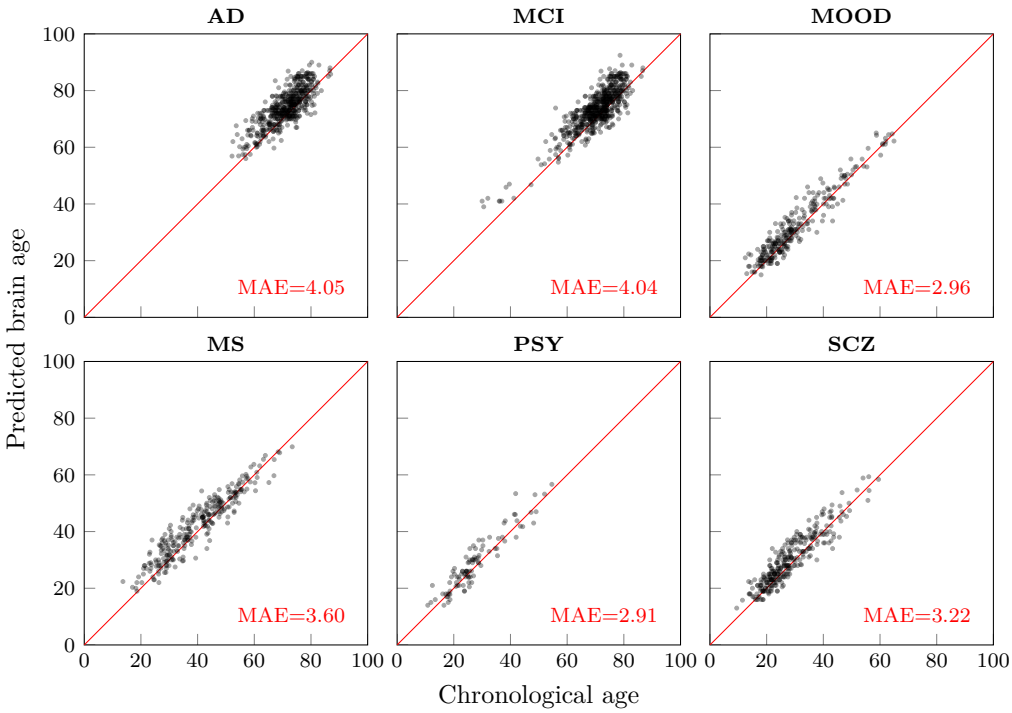


(c) Sources of generalization error

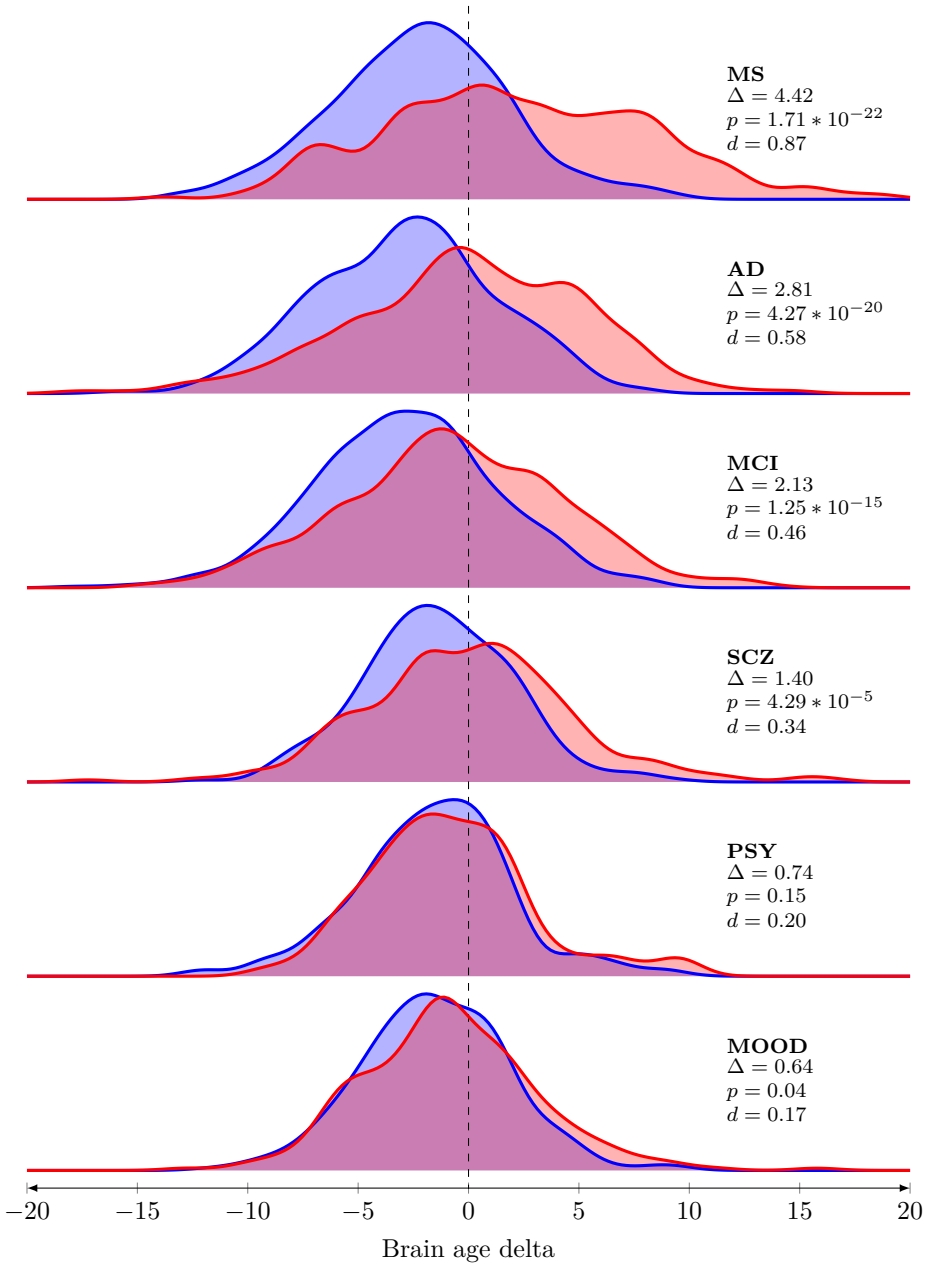
Supplementary Figure 1: The results of the model comparison on the resampled datasets representing a breakdown of the sources of generalization error. (a) For the "Unknown population" dataset we drew 100 bootstrapped subsets from the test set matching the empirical age and sex distributions of the external dataset, and tested the performance of SFCN-reg. (b) For the "Unknown scanners" dataset we similarily drew 100 bootstrapped subsets from the external dataset, matching the distributions of the test set.(c) A conceptualization of the assumed sources underlying the generalization error and the characteristics of the various datasets.



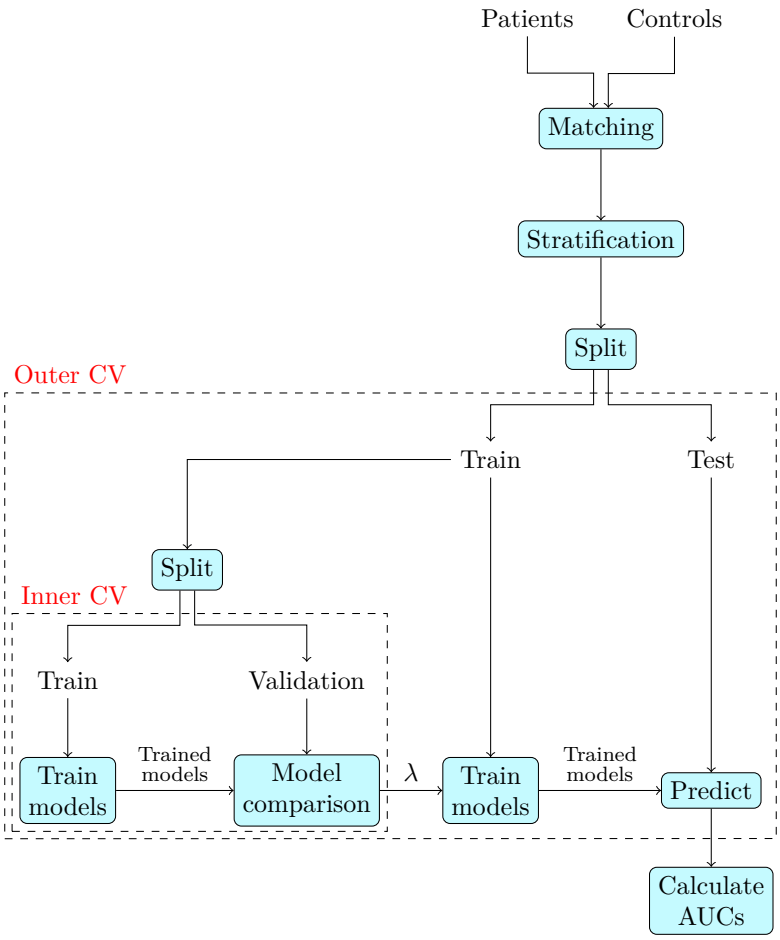
Supplementary Figure 2: Predictions of SFCN-reg on the external dataset, separated into the different scanning sites. Due to the large number of scanners, AIBL and ADNI is shown in single plots.



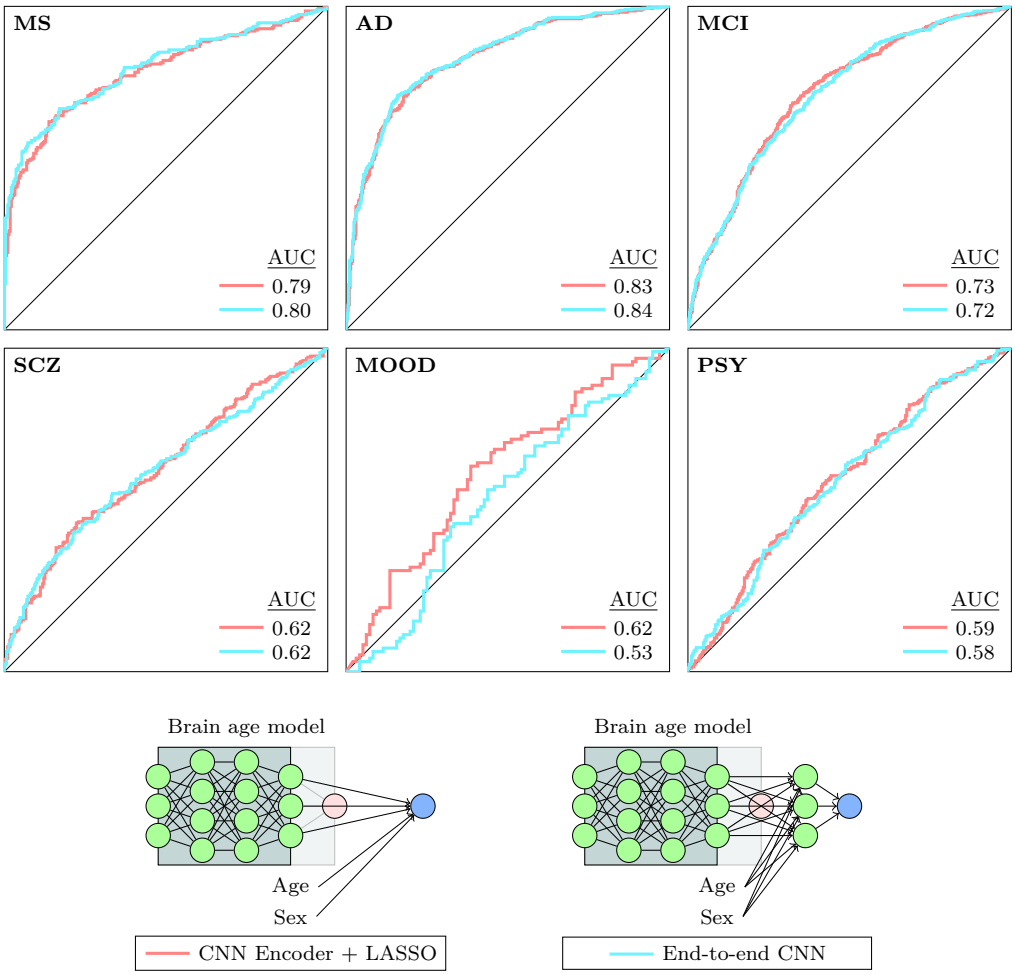
Supplementary Figure 3: Predictions of SFCN-reg on the control sets used in the clinical classifiers.



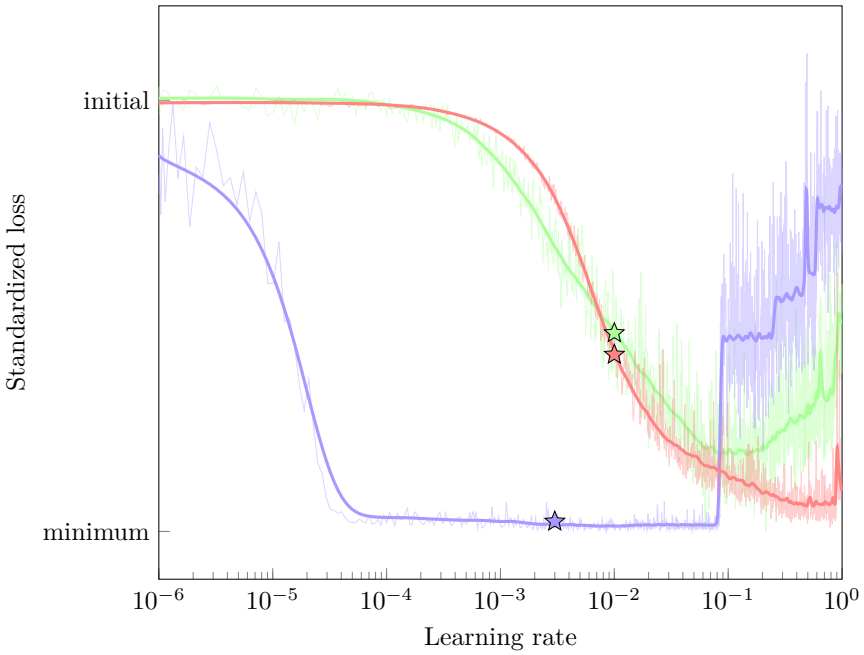
Supplementary Figure 4: The brain age delta distributions for the individual patient cohorts versus their individually matched control groups.



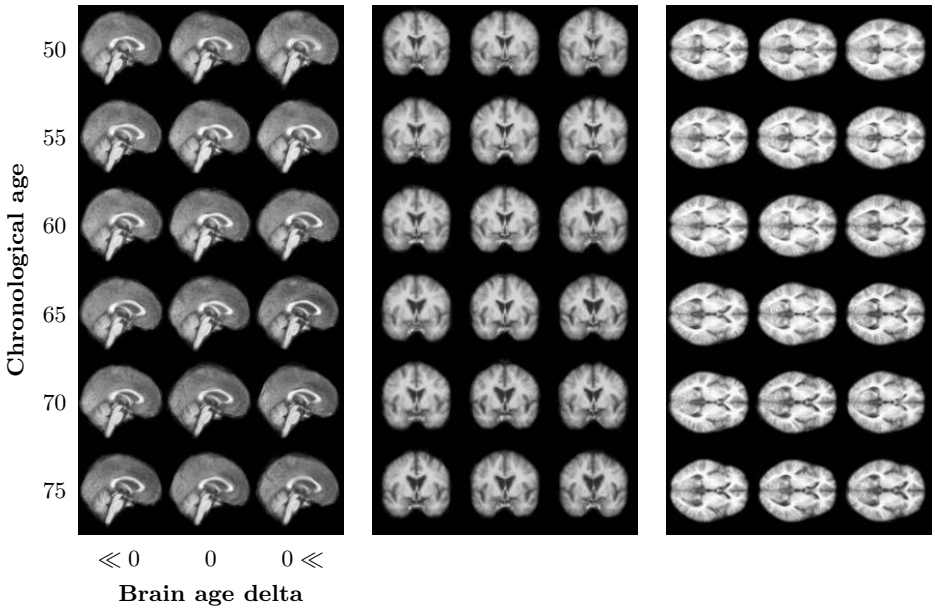
Supplementary Figure 5: The nested cross-validation (CV) scheme used for the clinical predictors. The outer CV ensures that the predictions used for the AUC calculations are on unseen data, while the inner loop optimizes the l_1-penalty parameter λ. On each level the data was split into 5 stratified folds, and the same folds were used across the three different LASSO models.



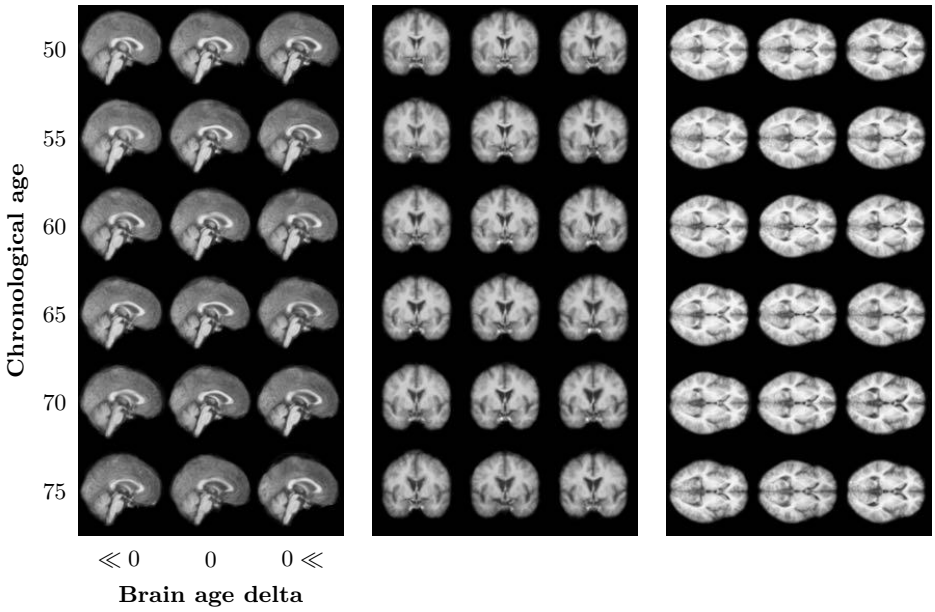
Supplementary Figure 6: A comparison of the AUCs of the LASSO and CNN clinical predictors. While both models rely on the features of the brain age model, the LASSO requires a two-step modelling process whereas the End-to-end CNN can be implemented as a single Keras model.



Supplementary Figure 7: Results of the learning rate sweep for the three model variants. The sweep investigated the models ability to learn at different learning rates, where the optimal range is indicated by a steep descent in the curve. The initial learning rates, denoted by the stars, were set manually based on this visualization.

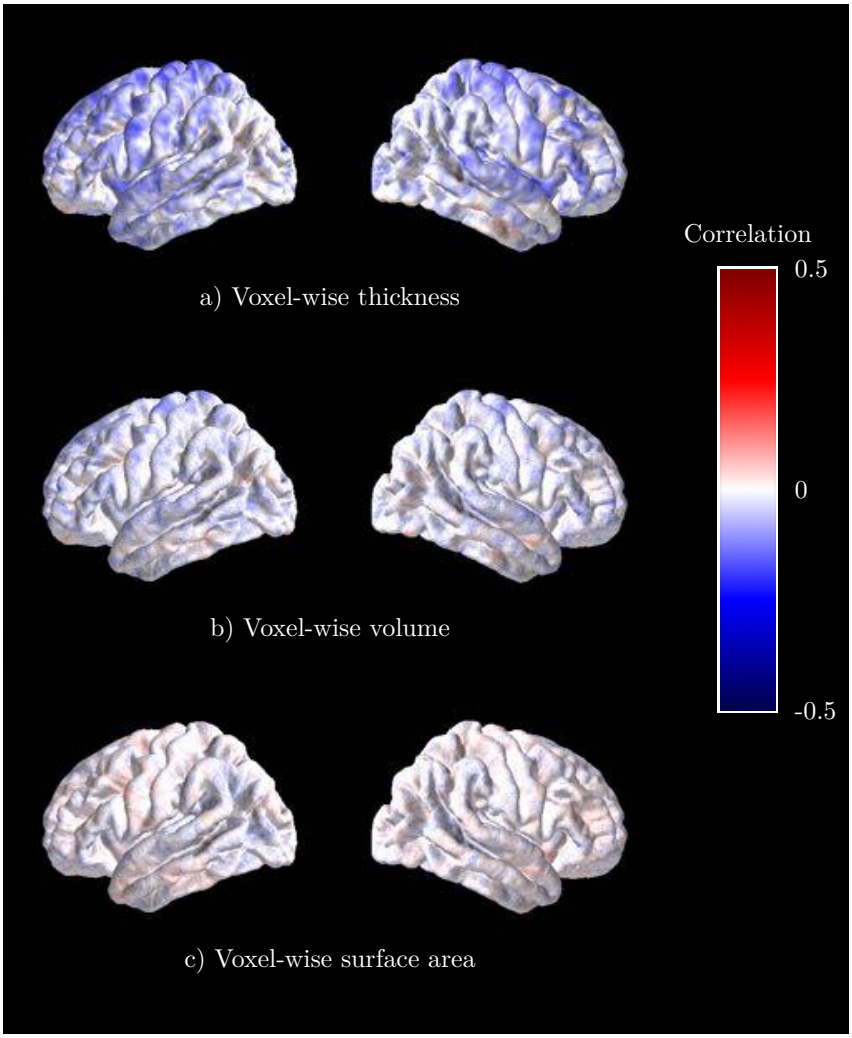


(a) Males



(b) Females

Supplementary Figure 8: Average brains for different chronological and brain age groups. Each row represents a single chronological age (+/- 0.5 years). Each column represents a different brain age delta category. The plotted brains are voxel-wise averages of the 10 participants making up the group.



Supplementary Figure 9: Correlation between brain age delta and cortical measures. Pearson correlation was computed with voxel-wise thickness (a), volume (b), and surface area (c), plotted on generic surfaces. The correlations were calculated using data from the largest scanner in the external dataset, the Oslo GE 750 scanner (n=876).

Name	Size	Subset	Ages	Sex (M/F)	Train	Validation	Test
Autism Brain Imaging Data Exchange	555	Healthy Controls	6-56	459/96	356	88	111
Autism Brain Imaging Data Exchange II	184	Healthy Controls	8-64	114/70	118	29	37
ADHD200	606	Healthy Controls	7-26	313/293	388	97	121
Amsterdam Open MRI Collection	819	ID1000	19-26	390/429	524	131	164
Beijing Normal University Enhanced Sample	180	Full	17-28	73/107	116	28	36
Cambridge Center for Aging and Neuroscience	653	Full	18-89	323/330	418	104	131
Consortium for Reliability and Reproducibility	1368	Full	6-84	672/696	880	219	269
Dallas Lifespan Brain Study	314	Full	21-89	117/197	201	50	63
ds000119	73	Full	8-27	30/43	47	11	15
ds000202	95	Full	18-30	0/95	61	15	19
ds000222	79	Full	21-73	38/41	51	12	16
1000 Functional Connectomes Project	812	Full	8-78	377/435	520	130	162
Healthy Brain Network	1855	Full	5-22	1183/672	1188	296	371
Human Connectome Project	1113	Full	22-37	512/601	712	178	223
Max Planck Institute Leipzig Mind-Brain-Body	73	Full	22-68	66/7	47	11	15
Enhanced Nathan Kline Institute - Rockland Sample	928	Full	6-85	363/565	594	148	186
Open Access Series of Imaging Studies	1264	Healthy Controls	42-95	495/769	817	188	259
Pediactric, Imaging and Neurocognition	1174	Full	3-21	610/564	752	187	235
Southwest University Adult Lifespan Dataset	494	Full	19-80	187/307	316	79	99
Southwest University Longitudinal Imaging Multimodal	573	Full	17-27	254/319	367	91	115
UK Biobank	40330	Custom*	45-82	19251/21079	25812	6452	8066
Total	**53542**	**-**	**3-95**	**25827/27715**	**34285**	**8544**	**10713**

Supplementary Table 1: An overview of the reference dataset. The dataset was compiled from 21 sources, spanning a wide age range and a multitude of scanners and scanning protocols. How many participants from each source was used for training, validation and testing is indicated in the respective columns.

Name	Size	Ages	Sex (M/F)
Alzheimer's Disease Neuroimaging Initiative	416	58-96	209/207
The Australian Imaging, Behaviour and Lifestyle Flagship Study	432	60-92	187/245
IXI	308	19-86	129/179
The Norwegian Cognitive NeuroGenetics Sample	209	23-81	72/137
StrokeMRI	341	20-94	128/213
TOP	848	13-72	445/390
Total	**2554**	**13-96**	**1170/1371**

Supplementary Table 2: An overview of the external dataset used for generalization testing. The external dataset only contains healthy controls and spans an age range slightly different than that of the reference dataset.

Cohort	Size	Ages	Source(s)
Alzheimer's disease	497	54-93	ADNI (338), AIBL (69), Demgen (116)
Mild cognitive impairment	603	38-92	ADNI (402), AIBL (85), Demgen (116)
Mixed psychotic diagnoses	96	14-57	TOP
Mood disorders	269	14-65	TOP
Multiple sclerosis	275	19-70	Multiple sclerosis
Schizophrenia	287	13-59	TOP

Supplementary Table 3: An overview of the patient cohorts used in the clinical classifiers. Each patient cohort has a matched control group, sampled from the same scanners with a resembling age and sex distribution.

Dataset	Exclusion criteria
ABIDE	Patient cohort Timepoints 2+ for longitudinal samples
ADHD200	Patient cohort
AOMIC	Timepoints 2+
FCON1000	Coming from ADHD cohort Beijing Zang cohort removed from test set
OASIS3	Patient cohort Timepoints 2+ for longitudinal samples
SLIM	Timepoints 2+ for longitudinal samples
UKBB	Having one of the following ICD-10 diagnoses: C70, C71, F[0,99], G[0-99],R[40-47],R90,S06 All timepoints for all longitudinal samples

Supplementary Table 4: Exclusion criteria for the participants in the reference dataset.

Dataset	Source	Comment	Reference(s)
Autism Brain Imaging Dataset Exchange	http://fcon_1000.projects.nitrc.org/	Primary support for the work by Adriana Di Martino was provided by the NIMH (K23MH087770) and the Leon Levy Foundation. Primary support for the work by Michael P. Milham and the INDI team was provided by gifts from Joseph P. Healy and the Stavros Niarchos Foundation to the Child Mind Institute, as well as by an NIMH award to MPM (R03MH096321).	[1]
Autism Brain Imaging Dataset Exchange II	http://fcon_1000.projects.nitrc.org/	Primary support for the work by Adriana Di Martino and her team was provided by the National Institute of Mental Health (NIMH 5R21MH107045). Primary support for the work by Michael P. Milham and his team provided by the National Institute of Mental Health (NIMH 5R21MH107045); Nathan S. Kline Institute of Psychiatric Research). Additional Support was provided by gifts from Joseph P. Healey, Phyllis Green and Randolph Cowen to the Child Mind Institute.	[2]
ADHD200	http://fcon_1000.projects.nitrc.org/	F. Xavier Castellanos, David Kennedy, Michael Milham, and Stewart Mostofsky are responsible for the initial conception of the ADHD-200 Consortium. Consortium steering committee includes Jan Buitelaar, F. Xavier Castellanos, Dan Dickstein, Damien Fair, David Kennedy, Beatriz Luna, Michael Milham (Project Coordinator), Stewart Mostofsky, and Julie Schweitzer. Data aggregation and organization was coordinated by the INDI team, which included Saroja Bangaru, David Gutman, Maarten Mennes, and Michael Milham. Web infrastructure and data storage were coordinated by Robert Buccigrossi, Albert Crowley, Christian Hasselgrove, David Kennedy, Kimberly Pohland, and Nina Preuss. The ADHD-200 Global Competition Coordinators were Damien Fair (Chair of Selection Committee, Editor in Chief for Global Competition Special issue) and Michael Milham	[3, 4]
Alzheimer's Disease Neuroimaging Initiative	http://adni.loni.usc.edu/	The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. ADNI consists of 4 waves, the later is still ongoing (ADNI 3). A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Data collection and sharing for this project was funded by the ADNI (NIH Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada and the Saguenay Yout Study. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.	
The Australian Imaging, Behaviour and Lifestyle Flagship Study	https://aibl.csiro.au/	Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) was funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which was made available at the ADNI database (http://www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at http://www.aibl.csiro.au. Correspondence should be addressed to Christopher Rowe (email: christopher.rowe@austin.org.au).	
The Amsterdam Open MRI Collection	https://nilab-uva.github.io/AOMIC.github.io/	We thank all research assistants and students who helped collecting the data of the three projects, Jasper Wijnen and Marco Teunisse for advice and guidance with respect to anonymization and GDPR-related concerns, and Jos Bloemers, Sennay Ghebeab, Adriaan Tuiten, Joram van Driel, Christian Olivers, Ilja Sligte, Sara Jahfari, Guido van Wingen, and Suzanne Oosterwijk for help with designing the paradigms, Marcus Spaan for technical support, and Franklin Feingold and Joe Wexler for help with uploading the datasets to Openneuro.	[5]
Beijing Normal University Enhanced Sample	http://fcon_1000.projects.nitrc.org/	Financial support for the data used in this project was provided by a grant from the National Natural Science Foundation of China: 30770594 and a grant from the National High Technology Program of China (863): 2008AA02Z405.	[6, 7]
Cam-CAN	https://camcan-archive.mrc-cbu.cam.ac.uk/dataaccess/	Data collection and sharing for this project was provided by the Cambridge Centre for Ageing and Neuroscience (CamCAN). CamCAN funding was provided by the UK Biotechnology and Biological Sciences Research Council (grant number BB/H008217/1), together with support from the UK Medical Research Council and University of Cambridge, UK.	[8, 9]

Dataset	Source	Comment	Reference(s)
Consortium for Reliability and Reproducibility	http://fcon_1000.projects.nitrc.org/		[10]
Demgen	Authors	Supported by the Norwegian National Advisory Unit on Aging and Health	[11, 12]
Dallas Lifespan Brain Study	http://fcon_1000.projects.nitrc.org/		[13]
ds000119	https://openfmri.org/	Supported by the National Institutes of Mental Health (NIMH RO1 MH067924). Enami Yasui provided assistance with data collection	[14]
ds000202	https://openfmri.org/		[15, 16]
ds000222	https://openfmri.org/		[17]
1000 Functional Connectomes Classic Sample	http://fcon_1000.projects.nitrc.org/	Collected at 33 independent sites by J.J. Pekar, S.H. Mostofsky, S, Colcombe, Y.F. Zang, D. Marguiles, R.L. Buckner, M.J Low, B. Rypma, D.J. Madden, A.C. Evans, S.A.R.B. Rombouts, A. Villringer, S.J. Li, C.Sorg, V. Riedel, B. Biswal, M. Hampson, M.P. Milham, F.X. Castellanos, P. Williamson, M. Hoptman, V.J. Kiviniemi, J. Veijiola, S.M. Smith, C. Mackay, M. Greicius, G. Siegle, K. McMahon, B. Schlaggar, S. Petersen, C.P. Lin, H.S. Mayberg, C.S. Monk, R.D. Seidler, S.J. Peltier	
Human Connectome Project	https://www.humanconnectome.org/	Data were provided [in part] by the Human Connectome Project, MGH-USC Consortium (PIs: Bruce R. Rosen, Arthur W. Toga and Van Wedeen; U01MH093765) funded by the NIH Blueprint Initiative for Neuroscience Research grant; the National Institutes of Health grant P41EB015896; and the Instrumentation Grants S10RR023043, 1S10RR023401, 1S10RR019307.	[18]
IXI	http://brain-development.org/ixi-dataset/		
Max Planck Institut Leipzig Mind-Brain-Body Dataset	http://fcon_1000.projects.nitrc.org/		[19]
Multiple sclerosis	Authors	The project was supported by grants from The Research Council of Norway (NFR, grant number 240102, 249795, and 223273) and the South-Eastern Health Authorities of Norway (grant number 257955, 2014097).	[20]
The Norwegian Cognitive NeuroGenetics Sample (NCNG)	Authors	Supported by grants from the the Research Council of Norway (177458/V50 and 231286/F20).	[21]
Enhanced Nathan Kline Institute - Rockland Sample	http://fcon_1000.projects.nitrc.org/	Principal support for the enhanced NKI-RS project is provided by the NIMH BRAINS R01MH094639-01 (PI Milham). Funding for key personnel also provided in part by the New York State Office of Mental Health and Research Foundation for Mental Hygiene. Funding for the decompression and augmentation of administrative and phenotypic protocols provided by a grant from the Child Mind Institute (1FDN2012-1). Additional personnel support provided by the Center for the Developing Brain at the Child Mind Institute, as well as NIMH R01MH081218, R01MH083246, and R21MH084126. Project support also provided by the NKI Center for Advanced Brain Imaging (CABI), the Brain Research Foundation, and the Stavros Niarchos Foundation.	[22]
Open Access Series of Imaging Studies 3	http://www.oasis-brains.org/	Supported by grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.	[23, 24]
Pediatric Imaging, Neurocognition and Genetics	http://pingstudy.ucsd.edu/	Data used in the preparation of this article were obtained from 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)). PING was a multisite, cross-sectional study that recruited more than 1,700 participants aged 3 to 20 years. The study was supported by award number RC2DA029475 from the National Institute on Drug Abuse with additional support for data sharing provided by the Eunice Kennedy Shriver National Institute of Child Health & Human Development under award number R01HD061414. A list of participating sites and study investigators can be found at https://ping-dataportal.ucsd.edu/sharing/Authors10222012.pdf. PING investigators designed and implemented the study and/or provided data but did not necessarily participate in 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.	[25]

Dataset	Source	Comment	Reference(s)
Southwest University Adult Lifespan Dataset	http://fcon_1000.projects.nitrc.org/		[26]
Southwest University Longitudinal Imaging Multimodal Brain Data	http://fcon_1000.projects.nitrc.org/	Support was provided by grant numbers 31271087; 31470981; 31571137, 31500885, SWU1509383, SWU1509451, cstc2015jcyjA10106, 151023, 2015M572423, 2015M580767, Xm2015037, 14JJD880009.	[27, 28]
StrokeMRI	Authors	Supported by the Research Council of Norway (249795, 248238), the South-Eastern Norway Regional Health Authority (2014097, 2015044, 2015073, 2016083), and the Norwegian ExtraFoundation for Health and Rehabilitation (2015/FO5146)	[29]
TOP	Authors	Supported by several grants from the Research Council of Norway, and the South-Eastern Norway Regional Health Authority	[30–32]
UK BioBank	https://www.ukbiobank.ac.uk/	This research has been conducted using the UK Biobank Resource (access code 27412).	[33, 34]

Supplementary Table 5: An overview of the sources of all the datasets used in the study.

Dataset	Number of scanners/protocols	Parameters
ABIDE	20	http://fcon_1000.projects.nitrc.org/indi/abide/scan_params/
ABIDE II	16	http://fcon_1000.projects.nitrc.org/indi/abide/scan_params/
ADHD200	6	Philips 1.5 T Gyroscan: TR=8ms, TE=3.76ms, FA=8° Siemens 3T Allegra: TR=2530ms, TE=3.25ms, FA=8° Siemens 3T Trio: TR=2300ms, TE=3.58ms, FA=10° Siemens 3T Trio: TR=1700ms, TE=3.92ms, FA=12° Siemens 3T Trio: TR=2100ms, TE=3.43ms, FA=8° Siemens 3T Trio: TR=2400ms, TE=3.08ms, FA=8°
ADNI	57	http://adni.loni.usc.edu/methods/mri-tool/mri-analysis/
AIBL	3	Siemens 1.5T Avanto: 3D MP-RAGE, TR = 2300ms, TE = 2.98ms, TI = 900, FA = 9° Siemens 3.0T Verio: 3D MP-RAGE, TR = 2300ms, TE = 2.98ms, TI = 900, FA = 9° Siemens 3.0T Tim Trio Siemens: 3D MP-RAGE, TR = 2300ms, TE = 2.98ms, TI = 900, FA = 9°
AOMIC	1	Philips 3T Intera: 3D MPRAGE TR=8.1ms, TE=3.7ms, FA=8°
Beijing Enhanced	1	SIEMENS TRIO 3T: MPRAGE,TR=2530ms,TE=3.39ms,FA=7°
Cam-CAN	1	Siemens 3T Trio: TR=2250ms, TE=2.99ms, FA=9°
CoRR	34	http://fcon_1000.projects.nitrc.org/indi/CoRR/html/_static/scan_parameters/
Demgen	2	GE 3T Signa HDxT: TR=7.8ms, TE=2.956ms, FA=12° GE 3T Discovery GE750: TR=8.16ms, TE=3.18ms, FA=12°
DLBS	1	Philips 3T: TR=8.135ms, TE=3.7ms, FA=18°
ds000119	1	Siemens 3T Allegra: TR=1570ms, TE=3.04ms, FA=8°
ds00202	1	Philips 3T Achieva: TR=7.6ms, TE=3.7ms,, FA=8°
ds000222	1	Siemens 3T Trio: TR=1550ms, TE=2.34ms, FA=9°
FCON 1000	28	http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html
HCP	1	Customized 3T scanner: TR=2400ms, TE=2.14, FA=8°
IXI	3	Philips 3T: TR=9.6ms, TE=4.6ms, FA=8° Philips 1.5T: TR=9.8ms, TE=4.6ms, FA=8° GE 1.5T: TR=6.0ms, TE=2.5ms
MPI-LEMON	1	Siemens MAGNETOM Verio 3T: MPRAGE,TR=680ms, TE1=5.19ms,TE2=7.65ms,FA=60°
Multiple sclerosis	2	Siements 1.5T Avanto: TR=2400ms, TE=3.61ms, FA=8° GE 3T Discovery GE750: TR=8.16ms, TE=3.18ms, FA=12°
NCNG	1	Siements 1.5T Avanto: TR=2400ms, TE=3.61ms, FA=8°
NKI Rockland	2	Siemens MAGNETOM TrioTim 3T: TR=1900ms, TE=2.52ms, FA=9° Siemens MAGNETOM TrioTim 3T: TR=2600ms, TE=3.02ms, FA=8°
OASIS 3	1	Siemens 1.5T Vision: TR=9.7ms, TE=4ms, FA=10°
PING	11	http://pingstudy.ucsd.edu/resources/neuroimaging-cores.html
SALD	1	Siemens 3T Trio: TR=1900ms, TE=2.52ms, FA=9°
SLIM	1	Siemens 3T Trio: TR=1900ms, TE=2.52ms, FA=9°
StrokeMRI	2	GE 3T Signa HDxT: TR=7.8ms, TE=2.956ms, FA=12° GE 3T Discovery GE750: TR=8.16ms, TE=3.18ms, FA=12°
TOP	4	Siemens 1.5T Sonata: TR=2730ms, TE=3.93ms, FA=7° GE 3T Signa HDxT: TR=7.8ms, TE=2.956ms, FA=12° (one subset with HNS coil, one subset with 8HRBRAIN coil) GE 3T Discovery GE750: TR=8.16ms, TE=3.18ms, FA=12°
UK Biobank	3	Siemens 3T Skyra: TR=2000ms, TE=2.01ms, FA=8° (3 identical scanning sites)

Supplementary Table 6: An overview of the scanning protocols used in the datasets.

Category	Columns
Alcohol	'Alcohol drinker status', 'Alcohol intake frequency', 'Former alcohol drinker', 'Average monthly red wine intake', 'Average monthly champagne plus white wine intake', 'Average monthly beer plus cider intake', 'Average monthly spirits intake', 'Average monthly fortified wine intake', 'Average monthly intake of other alcoholic drinks', 'Average weekly red wine intake', 'Average weekly champagne plus white wine intake', 'Average weekly beer plus cider intake', 'Average weekly spirits intake', 'Average weekly fortified wine intake', 'Average weekly intake of other alcoholic drinks', 'Alcohol usually taken with meals', 'Alcohol intake versus 10 years previously', 'Reason for reducing amount of alcohol drunk', 'Reason former drinker stopped drinking alcohol'
Biochemical	'Basophill count', 'Basophill percentage', 'Eosinophill count', 'Eosinophill percentage', 'Haematocrit percentage', 'Haemoglobin concentration', 'High light scatter reticulocyte count', 'High light scatter reticulocyte percentage', 'Immature reticulocyte fraction', 'Lymphocyte count', 'Lymphocyte percentage', 'Mean corpuscular haemoglobin', 'Mean corpuscular haemoglobin concentration', 'Mean corpuscular volume', 'Mean platelet (thrombocyte) volume', 'Mean reticulocyte volume', 'Mean sphered cell volume', 'Monocyte count', 'Monocyte percentage', 'Neutrophill count', 'Neutrophill percentage', 'Nucleated red blood cell count', 'Nucleated red blood cell percentage', 'Platelet count', 'Platelet crit', 'Platelet distribution width', 'Red blood cell (erythrocyte) count', 'Red blood cell (erythrocyte) distribution width', 'Reticulocyte count', 'Reticulocyte percentage', 'White blood cell (leukocyte) count', 'Alanine aminotransferase', 'Albumin', 'Alkaline phosphatase', 'Apolipoprotein A', 'Apolipoprotein B', 'Aspartate aminotransferase', 'C-reactive protein', 'Calcium', 'Cholesterol', 'Creatinine', 'Cystatin C', 'Direct bilirubin', 'Gamma glutamyltransferase', 'Glucose', 'Glycated haemoglobin (HbA1c)', 'HDL cholesterol', 'IGF-1', 'LDL direct', 'Lipoprotein A', 'Oestradiol', 'Phosphate', 'Rheumatoid factor', 'SHBG', 'Testosterone', 'Total bilirubin', 'Total protein', 'Triglycerides', 'Urate', 'Urea', 'Vitamin D', '3-Hydroxybutyrate', 'Acetate', 'Acetoacetate', 'Acetone', 'Alanine', 'Albumin', 'Apolipoprotein A1', 'Apolipoprotein B', 'Average Diameter for HDL Particles', 'Average Diameter for LDL Particles', 'Average Diameter for VLDL Particles', 'Cholesterol in Chylomicrons and Extremely Large VLDL', 'Cholesterol in IDL', 'Cholesterol in Large HDL', 'Cholesterol in Large LDL', 'Cholesterol in Large VLDL', 'Cholesterol in Medium HDL', 'Cholesterol in Medium LDL', 'Cholesterol in Medium VLDL', 'Cholesterol in Small HDL', 'Cholesterol in Small LDL', 'Cholesterol in Small VLDL', 'Cholesterol in Very Large HDL', 'Cholesterol in Very Large VLDL', 'Cholesterol in Very Small VLDL', 'Cholesteryl Esters in Chylomicrons and Extremely Large VLDL', 'Cholesteryl Esters in HDL', 'Cholesteryl Esters in IDL', 'Cholesteryl Esters in LDL', 'Cholesteryl Esters in Large HDL', 'Cholesteryl Esters in Large LDL', 'Cholesteryl Esters in Large VLDL', 'Cholesteryl Esters in Medium HDL', 'Cholesteryl Esters in Medium LDL', 'Cholesteryl Esters in Medium VLDL', 'Cholesteryl Esters in Small HDL', 'Cholesteryl Esters in Small LDL', 'Cholesteryl Esters in Small VLDL', 'Cholesteryl Esters in VLDL', 'Cholesteryl Esters in Very Large HDL', 'Cholesteryl Esters in Very Large VLDL', 'Cholesteryl Esters in Very Small VLDL', 'Citrate', 'Clinical LDL Cholesterol', 'Concentration of Chylomicrons and Extremely Large VLDL Particles', 'Concentration of HDL Particles', 'Concentration of IDL Particles', 'Concentration of LDL Particles', 'Concentration of Large HDL Particles', 'Concentration of Large LDL Particles', 'Concentration of Large VLDL Particles', 'Concentration of Medium HDL Particles', 'Concentration of Medium LDL Particles', 'Concentration of Medium VLDL Particles', 'Concentration of Small HDL Particles', 'Concentration of Small LDL Particles', 'Concentration of Small VLDL Particles', 'Concentration of VLDL Particles', 'Concentration of Very Large HDL Particles', 'Concentration of Very Large VLDL Particles', 'Concentration of Very Small VLDL Particles', 'Creatinine', 'Degree of Unsaturation', 'Docosahexaenoic Acid', 'Free Cholesterol in Chylomicrons and Extremely Large VLDL', 'Free Cholesterol in HDL', 'Free Cholesterol in IDL', 'Free Cholesterol in LDL', 'Free Cholesterol in Large HDL', 'Free Cholesterol in Large LDL', 'Free Cholesterol in Large VLDL', 'Free Cholesterol in Medium HDL', 'Free Cholesterol in Medium LDL', 'Free Cholesterol in Medium VLDL', 'Free Cholesterol in Small HDL', 'Free Cholesterol in Small LDL', 'Free Cholesterol in Small VLDL', 'Free Cholesterol in VLDL', 'Free Cholesterol in Very Large HDL', 'Free Cholesterol in Very Large VLDL', 'Free Cholesterol in Very Small VLDL', 'Glucose', 'Glutamine', 'Glycine', 'Glycoprotein Acetyls', 'HDL Cholesterol', 'Histidine', 'Isoleucine', 'LDL Cholesterol', 'Lactate', 'Leucine', 'Linoleic Acid', 'Monounsaturated Fatty Acids', 'Omega-3 Fatty Acids', 'Omega-6 Fatty Acids', 'Phenylalanine', 'Phosphatidylcholines', 'Phosphoglycerides', 'Phospholipids in Chylomicrons and Extremely Large VLDL', 'Phospholipids in HDL', 'Phospholipids in IDL', 'Phospholipids in LDL', 'Phospholipids in Large HDL', 'Phospholipids in Large LDL', 'Phospholipids in Large VLDL', 'Phospholipids in Medium HDL', 'Phospholipids in Medium LDL', 'Phospholipids in Medium VLDL', 'Phospholipids in Small HDL', 'Phospholipids in Small LDL', 'Phospholipids in Small VLDL', 'Phospholipids in VLDL', 'Phospholipids in Very Large HDL', 'Phospholipids in Very Large VLDL', 'Phospholipids in Very Small VLDL', 'Polyunsaturated Fatty Acids', 'Pyruvate', 'Remnant Cholesterol (Non-HDL, Non-LDL -Cholesterol)', 'Saturated Fatty Acids', 'Sphingomyelins', 'Total Cholesterol', 'Total Cholesterol Minus HDL-C', 'Total Cholines', 'Total Concentration of Branched-Chain Amino Acids (Leucine + Isoleucine + Valine)', 'Total Concentration of Lipoprotein Particles', 'Total Esterified Cholesterol', 'Total Fatty Acids', 'Total Free Cholesterol', 'Total Lipids in Chylomicrons and Extremely Large VLDL', 'Total Lipids in HDL', 'Total Lipids in IDL', 'Total Lipids in LDL', 'Total Lipids in Large HDL', 'Total Lipids in Large LDL', 'Total Lipids in Large VLDL', 'Total Lipids in Lipoprotein Particles', 'Total Lipids in Medium HDL', 'Total Lipids in Medium LDL', 'Total Lipids in Medium VLDL', 'Total Lipids in Small HDL', 'Total Lipids in Small LDL', 'Total Lipids in Small VLDL', 'Total Lipids in VLDL', 'Total Lipids in Very Large HDL', 'Total Lipids in Very Large VLDL', 'Total Lipids in Very Small VLDL', 'Total Phospholipids in Lipoprotein Particles', 'Total Triglycerides', 'Triglycerides in Chylomicrons and Extremely Large VLDL', 'Triglycerides in HDL', 'Triglycerides in IDL', 'Triglycerides in LDL', 'Triglycerides in Large HDL', 'Triglycerides in Large LDL', 'Triglycerides in Large VLDL', 'Triglycerides in Medium HDL', 'Triglycerides in Medium LDL', 'Triglycerides in Medium VLDL', 'Triglycerides in Small HDL', 'Triglycerides in Small LDL', 'Triglycerides in Small VLDL', 'Triglycerides in VLDL', 'Triglycerides in Very Large HDL', 'Triglycerides in Very Large VLDL', 'Triglycerides in Very Small VLDL', 'Tyrosine', 'VLDL Cholesterol', 'Valine', 'Creatinine (enzymatic) in urine', 'Microalbumin in urine', 'Potassium in urine', 'Sodium in urine'

Cognitive	'Mean time to correctly identify matches"', 'Digits entered correctly', 'Maximum digits remembered correctly', 'Time first key touched', 'Time last key touched', 'Time elapsed', 'Time to complete test', 'Fluid intelligence score', 'Number of fluid intelligence questions attempted within time limit', 'Attempted fluid intelligence (FI) test', 'FI1 : numeric addition test', 'FI2 : identify largest number', 'FI3 : word interpolation', 'FI4 : positional arithmetic', 'FI5 : family relationship calculation', 'FI6 : conditional arithmetic', 'FI7 : synonym', 'FI8 : chained arithmetic', 'FI9 : concept interpolation', 'FI10 : arithmetic sequence recognition', 'FI11 : antonym', 'FI12 : square sequence recognition', 'FI13 : subset inclusion logic', 'Duration to complete numeric path (trail #1)', 'Duration to complete alphanumeric path (trail #2)', 'Total errors traversing numeric path (trail #1)', 'Total errors traversing alphanumeric path (trail #2)', 'Interval between previous point and current one in numeric path (trail #1)', 'Interval between previous point and current one in alphanumeric path (trail #2)', 'Errors before selecting correct item in numeric path (trail #1)', 'Errors before selecting correct item in alphanumeric path (trail #2)', 'Number of puzzles correctly solved', 'Number of puzzles viewed', 'Duration spent answering each puzzle', 'Vocabulary level', 'Number of word pairs correctly associated', 'Word associated with huge', 'Word associated with happy', 'Word associated with tattered', 'Word associated with old', 'Word associated with long', 'Word associated with red', 'Word associated with sulking', 'Word associated with pretty', 'Word associated with tiny', 'Word associated with new', 'Prospective memory result', 'Number of incorrect matches in round', 'Time to complete round', 'Number of correct matches in round'
Diet	'Cooked vegetable intake', 'Salad / raw vegetable intake', 'Fresh fruit intake', 'Dried fruit intake', 'Oily fish intake', 'Non-oily fish intake', 'Processed meat intake', 'Poultry intake', 'Beef intake', 'Lamb/mutton intake', 'Pork intake', 'Age when last ate meat', 'Never eat eggs, dairy, wheat, sugar', 'Cheese intake', 'Milk type used', 'Spread type', 'Bread intake', 'Cereal intake', 'Salt added to food', 'Tea intake', 'Coffee intake', 'Water intake', 'Major dietary changes in the last 5 years', 'Variation in diet'
Early life factors	'Ease of skin tanning', 'Hair colour (natural, before greying)', 'Country of birth', 'Breastfed as a baby', 'Comparative body size at age 10', 'Comparative height size at age 10', 'Handedness', 'Adopted as a child', 'Part of a multiple birth', 'Relative age of first facial hair', 'Relative age voice broke', 'Birth weight', 'Country of Birth (non-UK origin)'
Family	'Illnesses of adopted father', 'Illnesses of adopted mother', 'Illnesses of adopted siblings', 'Illnesses of father', 'Illnesses of mother', 'Illnesses of siblings', 'Father still alive', "Father's age", "Father's age at death", 'Mother still alive', "Mother's age", "Mother's age at death", 'Number of full brothers', 'Number of adopted brothers', 'Number of full sisters', 'Number of adopted sisters', 'Number of older siblings', 'Non-accidental death in close genetic family'
Health related	'Facial ageing', 'Mouth/teeth dental problems', 'Overall health rating', 'Long-standing illness, disability or infirmity', 'Falls in the last year', 'Wheeze or whistling in the chest in last year', 'Shortness of breath walking on level ground', 'Pain type(s) experienced in last month', 'Headaches for 3+ months', 'Facial pains for 3+ months', 'Neck/shoulder pain for 3+ months', 'Back pain for 3+ months', 'Stomach/abdominal pain for 3+ months', 'General pain for 3+ months', 'Had major operations', 'Had other major operations', 'Vascular/heart problems diagnosed by doctor', 'Age heart attack diagnosed', 'Age angina diagnosed', 'Age stroke diagnosed', 'Age high blood pressure diagnosed', 'Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor', 'Age deep-vein thrombosis (DVT, blood clot in leg) diagnosed', 'Age pulmonary embolism (blood clot in lung) diagnosed', 'Age emphysema/chronic bronchitis diagnosed', 'Age asthma diagnosed', 'Age hay fever, rhinitis or eczema diagnosed', 'Diabetes diagnosed by doctor', 'Gestational diabetes only', 'Age diabetes diagnosed', 'Started insulin within one year diagnosis of diabetes', 'Cancer diagnosed by doctor', 'Fractured/broken bones in last 5 years', 'Fractured bone site(s)', 'Fracture resulting from simple fall', 'Other serious medical condition/disability diagnosed by doctor', 'Medication for cholesterol, blood pressure or diabetes', 'Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones', 'Taking other prescription medications', 'Medication for pain relief, constipation, heartburn', 'Vitamin and mineral supplements', 'Mineral and other dietary supplements', 'Hearing difficulty/problems', 'Hearing aid user', 'Cochlear implant', 'Tinnitus', 'Hair/balding pattern', 'Age when periods started (menarche)', 'Had menopause', 'Age at menopause (last menstrual period)', 'Time since last menstrual period', 'Length of menstrual cycle', 'Menstruating today', 'Number of live births', 'Age at first live birth', 'Age at last live birth', 'Ever had stillbirth, spontaneous miscarriage or termination', 'Ever taken oral contraceptive pill', 'Age started oral contraceptive pill', 'Ever had hysterectomy (womb removed)', 'Cancer code, self-reported', 'Cancer year/age first occurred', 'Pregnant', 'Pace-maker', 'Glasses worn/required (left)', 'Glasses worn/required (right)', 'Wears glasses or contact lenses', 'Age started wearing glasses or contact lenses', 'Reason for glasses/contact lenses', 'Other eye problems', 'Eye problems/disorders', 'Age when diabetes-related eye disease diagnosed', 'Age glaucoma diagnosed', 'Age when loss of vision due to injury or trauma diagnosed', 'Age cataract diagnosed', 'Age macular degeneration diagnosed', 'Age other serious eye condition diagnosed'
Lifestyle	'Length of mobile phone use', 'Weekly use of mobile phone in last 3 months', 'Plays computer games', 'Sleep duration', 'Getting up in morning', 'Morning/evening person', 'Nap during day', 'Sleeplessness/insomnia', 'Daytime dozing/sleeping (narcolepsy)', 'Time spend outdoors in summer', 'Time spent outdoors in winter', 'Use of sun/uv protection', 'Frequency of solarium/sunlamp use', 'Age first had sexual intercourse', 'Lifetime number of sexual partners', 'Ever had same-sex intercourse', 'Lifetime number of same-sex sexual partners', 'Frequency of friend/family visits', 'Leisure/social activities'
Mental health	'Able to confide', 'Bipolar and major depression status', 'Bipolar disorder status', 'Neuroticism score', 'Probable recurrent major depression (moderate)', 'Probable recurrent major depression (severe)', 'Single episode of probable major depression', 'Mood swings', 'Miserableness', 'Irritability', 'Sensitivity / hurt feelings', 'Fed-up feelings', 'Nervous feelings', 'Worrier / anxious feelings', "Tense / 'highly strung'", 'Worry too long after embarrassment', "Suffer from 'nerves'", 'Loneliness, isolation', 'Guilty feelings', 'Risk taking', 'Happiness', 'Work/job satisfaction', 'Health satisfaction', 'Family relationship satisfaction', 'Friendships satisfaction', 'Financial situation satisfaction', 'Seen doctor (GP) for nerves, anxiety, tension or depression', 'Seen a psychiatrist for nerves, anxiety, tension or depression'

	'Ever depressed for a whole week', 'Longest period of depression', 'Number of depression episodes', 'Ever unenthusiastic/disinterested for a whole week', 'Longest period of unenthusiasm / disinterest', 'Number of unenthusiastic/disinterested episodes', 'Ever manic/hyper for 2 days', 'Ever highly irritable/argumentative for 2 days', 'Manic/hyper symptoms', 'Length of longest manic/irritable episode', 'Severity of manic/irritable episodes', 'Illness, injury, bereavement, stress in last 2 years'
Microbial infection	'HSV-1 seropositivity for Herpes Simplex virus-1', 'HSV-2 seropositivity for Herpes Simplex virus-2', 'VZV seropositivity for Varicella Zoster Virus', 'EBV seropositivity for Epstein-Barr Virus', 'CMV seropositivity for Human Cytomegalovirus', 'HHV-6 overall seropositivity for Human Herpesvirus-6', 'HHV-6A seropositivity for Human Herpesvirus-6', 'HHV-6B seropositivity for Human Herpesvirus-6', 'HHV-7 seropositivity for Human Herpesvirus-7', "KSHV seropositivity for Kaposi's Sarcoma-Associated Herpesvirus", 'HBV seropositivity for Hepatitis B Virus', 'HCV seropositivity for Hepatitis C Virus', 'T. gondii seropositivity for Toxoplasma gondii', 'HTLV-1 seropositivity for Human T-Lymphotropic Virus 1', 'HIV-1 seropositivity for Human Immunodeficiency Virus', 'BKV seropositivity for Human Polyomavirus BKV', 'JCV seropositivity for Human Polyomavirus JCV', 'MCV seropositivity for Merkel Cell Polyomavirus', 'HPV 16 Definition I seropositivity for Human Papillomavirus type-16', 'HPV 16 Definition II seropositivity for Human Papillomavirus type-16', 'HPV 18 seropositivity for Human Papillomavirus type-18', 'C. trachomatis Definition I seropositivity for Chlamydia trachomatis', 'C. trachomatis Definition II seropositivity for Chlamydia trachomatis', 'H. pylori Definition I seropositivity for Helicobacter pylori', 'H. pylori Definition II seropositivity for Helicobacter pylori'
Physical measures	'Diastolic blood pressure, automated reading', 'Diastolic blood pressure, manual reading', 'Pulse rate (during blood-pressure measurement)', 'Pulse rate, automated reading', 'Systolic blood pressure, automated reading', 'Systolic blood pressure, manual reading', 'Absence of notch position in the pulse waveform', 'Arterial pulse-wave stiffness device ID', 'Arterial stiffness device ID', 'Position of pulse wave notch', 'Position of the pulse wave peak', 'Position of the shoulder on the pulse waveform', 'Pulse rate', 'Pulse wave Arterial Stiffness index', 'Pulse wave peak to peak time', 'Pulse wave reflection index', 'Pulse wave velocity (manual entry)', 'Hand grip strength (left)', 'Hand grip strength (right)', 'Height', 'Weight (pre-imaging)', 'Waist circumference', 'Weight', 'Body mass index (BMI)', 'Hip circumference', 'Basal metabolic rate', 'Body fat percentage', 'Ventricular rate'
Smoking	'Ever smoked', 'Pack years adult smoking as proportion of life span exposed to smoking', 'Ever smoked', 'Pack years adult smoking as proportion of life span exposed to smoking', 'Pack years of smoking', 'Smoking status', 'Current tobacco smoking', 'Past tobacco smoking', 'Light smokers, at least 100 smokes in lifetime', 'Age started smoking in current smokers', 'Type of tobacco currently smoked', 'Previously smoked cigarettes on most/all days', 'Number of cigarettes currently smoked daily (current cigarette smokers)', 'Age stopped smoking cigarettes (current cigar/pipe or previous cigarette smoker)', 'Number of cigarettes previously smoked daily (current cigar/pipe smokers)', 'Time from waking to first cigarette', 'Difficulty not smoking for 1 day', 'Ever tried to stop smoking', 'Wants to stop smoking', 'Smoking compared to 10 years previous', 'Age started smoking in former smokers', 'Type of tobacco previously smoked', 'Number of cigarettes previously smoked daily', 'Age stopped smoking', 'Ever stopped smoking for 6+ months', 'Number of unsuccessful stop-smoking attempts', 'Likelihood of resuming smoking', 'Smoking/smokers in household', 'Exposure to tobacco smoke at home', 'Exposure to tobacco smoke outside home', 'Maternal smoking around birth'
Socioeconomic	'Number in household', 'Number of vehicles in household', 'Average total household income before tax', 'Current employment status', 'Job involves mainly walking or standing', 'Job involves heavy manual or physical work', 'Qualifications (Education)', 'Age completed full time education', 'Attendance/disability/mobility allowance', 'Private healthcare'

Supplementary Table 7: The columns used in the PheWAS, manually divided into the thirteen categories.

Loss function	Dropout	Weight decay	MAE_{TRAIN}	MAE_{VAL}
Soft Classification	0.5	1e-3	1.79	2.34
Soft Classification	**0.3**	**1e-3**	**1.57**	**2.28**
Soft Classification	0.3	1e-4	1.99	2.42
Regression	0.3	1e-3	1.96	2.67
Regression	**0.2**	**1e-4**	**1.67**	**2.51**
Regression	0.0	0.0	1.26	2.53
Ranking	**0.3**	**1e-3**	**2.38**	**2.59**
Ranking	0.2	1e-4	2.95	3.03
Ranking	0.5	1e-3	2.53	2.70

Supplementary Table 8: The hyperparameter settings tested for each of the three model variants. The best version for each model variant was selected based on the Mean Absolute Error in the validation set (MAE_{VAL}). The hyperparameters were determined sequentially, based on the performance of the previous runs.

Model	MAE_{TEST}	$\text{MAE}_{\text{EXTERNAL}}$	Generalization gap $\left(\frac{MAE_{EXTERNAL}}{MAE_{TEST}}\right)$
SFCN-sm	**2.23**	5.04	2.26
SFCN-reg	2.47	**3.90**	**1.57**
SFCN-rank	2.55	5.92	2.32

Supplementary Table 9: Results of the final model comparison. The test Mean Absolute Error (MAE_{TEST}) was computed from the test set, a portion of the reference dataset unseen during optimization. The external Mean Absolute Error ($\text{MAE}_{\text{EXTERNAL}}$) was computed in the external dataset, comprised of different origins, containing other scanners and a diverging age distribution compared to that of the training set. The generalization gap indicates the expected degree of deterioration when transferring to a new dataset.

Dataset	MAE	Mean	Standard deviation	Slope
Test split	2.47	-0.02	3.16	-0.03
External dataset	3.90	-2.13	4.63	-0.06

Supplementary Table 10: Model performance in the two datasets used in the final model comparison. Both the mean and standard references the brain age delta in the respective dataset, representing the bias and precision of the predictions. The slope is the regression coefficient of a linear model trained to predict delta from chronological age, representing the angle of the regression line compared to a perfect model.

Structure	Correlation
Right-Hippocampus	-0.20
3rd-Ventricle	0.18
Right-Accumbens-area	-0.18
Left-Hippocampus	-0.17
Left-Amygdala	-0.16
Right-Amygdala	-0.15
Left-Inf-Lat-Vent	0.15
CSF	0.15
Right-Lateral-Ventricle	0.14
non-WM-hypointensities	0.13
WM-hypointensities	0.13
Left-Lateral-Ventricle	0.13
Left-Thalamus-Proper	-0.12
Right-Inf-Lat-Vent	0.12
Right-Thalamus-Proper	-0.09
CC_Central	-0.09
CC_Anterior	-0.09
Left-Pallidum	0.08
CC_Posterior	-0.08
CC_Mid_Anterior	-0.08
Right-VentralDC	-0.07
Right-Cerebellum-Cortex	-0.07
Left-VentralDC	-0.07
Right-choroid-plexus	0.07
Left-Cerebellum-White-Matter	-0.06
Left-Putamen	-0.06
Left-Cerebellum-Cortex	-0.06
Optic-Chiasm	0.06
Right-Cerebellum-White-Matter	-0.06
CC_Mid_Posterior	-0.06
Left-Accumbens-area	-0.04
Right-Putamen	-0.04
4th-Ventricle	0.04
Left-choroid-plexus	0.04
Right-vessel	0.03
Brain-Stem	-0.03
Right-Caudate	0.02
Right-Pallidum	0.02

Structure	Correlation
5th-Ventricle	-0.01
Left-Caudate	0.01
Left-vessel	0.00

Supplementary Table 11: Correlation between subcortical volumes and the brain age delta. The subcortical volumes were computed using FreeSurfer's recon-all pipeline in the portion of the external dataset coming from the Oslo GE 750 scanner. Then we computed the Pearson correlation univariately between each volume and the brain age delta originating from SFCN-reg.

Supplementary References

1. Di Martino, A. *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry* **19,** 659–667 (2014).
2. Di Martino, A. *et al.* Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Scientific data* **4,** 1–15 (2017).
3. Brown, M. R. *et al.* ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Frontiers in systems neuroscience* **6,** 69 (2012).
4. Milham, M. P., Fair, D., Mennes, M., Mostofsky, S. H., *et al.* The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience* **6,** 62 (2012).
5. Snoek, L. *et al.* The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. *Scientific data* **8,** 1–23 (2021).
6. Yan, C. & Zang, Y. DPARSF: a MATLAB toolbox for" pipeline" data analysis of resting-state fMRI. *Frontiers in systems neuroscience* **4,** 13 (2010).
7. Tian, L., Wang, J., Yan, C. & He, Y. Hemisphere-and gender-related differences in small-world brain networks: a resting-state functional MRI study. *Neuroimage* **54,** 191–202 (2011).
8. Shafto, M. A. *et al.* The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC neurology* **14,** 1–25 (2014).
9. Taylor, J. R. *et al.* The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *Neuroimage* **144,** 262–269 (2017).
10. Zuo, X.-N. *et al.* An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data* **1,** 1–13 (2014).
11. Córdova-Palomera, A. *et al.* Disrupted global metastability and static and dynamic brain connectivity across individuals in the Alzheimer's disease continuum. eng. *Scientific Reports* **7,** 40268. ISSN: 2045-2322 (Jan. 2017).
12. Doan, N. T. *et al.* Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum. eng. *Scientific Reports* **7,** 45131. ISSN: 2045-2322 (Mar. 2017).
13. Lu, H. *et al.* Alterations in cerebral metabolic rate and blood supply across the adult lifespan. *Cerebral cortex* **21,** 1426–1434 (2011).
14. Velanova, K., Wheeler, M. E. & Luna, B. Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cerebral cortex* **18,** 2505–2522 (2008).
15. Van Schuerbeek, P., Baeken, C., De Raedt, R., De Mey, J. & Luypaert, R. Individual differences in local gray and white matter volumes reflect differences in temperament and character: a voxel-based morphometry study in healthy young females. *Brain Research* **1371,** 32–42 (2011).

16. Van Schuerbeek, P., Baeken, C. & De Mey, J. The heterogeneity in retrieved relations between the personality trait 'harm avoidance'and gray matter volumes due to variations in the VBM and ROI labeling processing settings. *PloS one* **11,** e0153865 (2016).

17. FitzGerald, T. H., Hämmerer, D., Friston, K. J., Li, S.-C. & Dolan, R. J. Sequential inference as a mode of cognition and its correlates in fronto-parietal and hippocampal brain regions. *PLoS computational biology* **13,** e1005418 (2017).

18. Van Essen, D. C. *et al.* The WU-Minn human connectome project: an overview. *Neuroimage* **80,** 62–79 (2013).

19. Babayan, A. *et al.* A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Scientific data* **6,** 1–21 (2019).

20. Høgestøl, E. A. *et al.* Cross-Sectional and Longitudinal MRI Brain Scans Reveal Accelerated Brain Aging in Multiple Sclerosis. *Frontiers in Neurology* **10,** 450. ISSN: 1664-2295 (2019).

21. Espeseth, T. *et al.* Imaging and cognitive genetics: the Norwegian Cognitive NeuroGenetics sample. *Twin Research and Human Genetics* **15,** 442–452 (2012).

22. Nooner, K. B. *et al.* The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Frontiers in neuroscience* **6,** 152 (2012).

23. Buckner, R. L. *et al.* A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* **23,** 724–738 (2004).

24. Fotenos, A. F., Snyder, A., Girton, L., Morris, J. & Buckner, R. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology* **64,** 1032–1039 (2005).

25. Jernigan, T. L. *et al.* The pediatric imaging, neurocognition, and genetics (PING) data repository. *Neuroimage* **124,** 1149–1154 (2016).

26. Wei, D. *et al.* Structural and functional brain scans from the cross-sectional Southwest University adult lifespan dataset. *Scientific data* **5,** 1–10 (2018).

27. Wang, Y., Wei, D., Li, W. & Qiu, J. Individual differences in brain structure and resting-state functional connectivity associated with type A behavior pattern. *Neuroscience* **272,** 217–228 (2014).

28. Zhu, W. *et al.* Brain structure links everyday creativity to creative achievement. *Brain and Cognition* **103,** 70–76 (2016).

29. Dørum, E. S. *et al.* Age-related differences in brain network activation and coactivation during multiple object tracking. *Brain and behavior* **6,** e00533 (2016).

30. Rimol, L. M. *et al.* Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. eng. *Biological Psychiatry* **68,** 41–50. ISSN: 1873-2402 (July 2010).

31. Tønnesen, S. *et al.* White matter aberrations and age-related trajectories in patients with schizophrenia and bipolar disorder revealed by diffusion tensor imaging. eng. *Scientific Reports* **8,** 14129. ISSN: 2045-2322 (Sept. 2018).

32. Doan, N. T. *et al.* Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. eng. *NeuroImage. Clinical* **15,** 719–731. ISSN: 2213-1582 (2017).

33. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12,** e1001779 (2015).

34. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018).

II

Molecular Psychiatry www.nature.com/mp

ARTICLE OPEN

Check for updates

Genetic architecture of brain age and its causal relations with brain and mental disorders

Esten H. Leonardsen[1,2], Didac Vidal-Piñeiro[1], James M. Roe[1], Oleksandr Frei[2], Alexey A. Shadrin[2,3], Olena Iakunchykova[1], Ann-Marie G. de Lange[4,5,6], Tobias Kaufmann[2,7], Bernd Taschler[8], Stephen M. Smith[8], Ole A. Andreassen[2,3], Thomas Wolfers[2,4,7], Lars T. Westlye[2,3,4] and Yunpeng Wang[1]✉

The difference between chronological age and the apparent age of the brain estimated from brain imaging data—the brain age gap (BAG)—is widely considered a general indicator of brain health. Converging evidence supports that BAG is sensitive to an array of genetic and nongenetic traits and diseases, yet few studies have examined the genetic architecture and its corresponding causal relationships with common brain disorders. Here, we estimate BAG using state-of-the-art neural networks trained on brain scans from 53,542 individuals (age range 3–95 years). A genome-wide association analysis across 28,104 individuals (40–84 years) from the UK Biobank revealed eight independent genomic regions significantly associated with BAG ($p < 5 \times 10^{-8}$) implicating neurological, metabolic, and immunological pathways – among which seven are novel. No significant genetic correlations or causal relationships with BAG were found for Parkinson's disease, major depressive disorder, or schizophrenia, but two-sample Mendelian randomization indicated a causal influence of AD ($p = 7.9 \times 10^{-4}$) and bipolar disorder ($p = 1.35 \times 10^{-2}$) on BAG. These results emphasize the polygenic architecture of brain age and provide insights into the causal relationship between selected neurological and neuropsychiatric disorders and BAG.

Molecular Psychiatry (2023) 28:3111–3120; https://doi.org/10.1038/s41380-023-02087-y

INTRODUCTION

Over the last decade, brain age has emerged as a promising measure of overall brain health [1, 2]. To estimate brain age, machine learning models are applied to brain imaging data to learn visual patterns characteristic of different ages [3, 4]. The difference between predicted brain age and chronological age is termed the brain age gap (BAG) and indicates deviation from a normative trajectory, a potential health indicator. Earlier studies have found a large variation in the predicted brain age of individuals with the same chronological age, and that these interindividual variations correlate with neurological and mental disorders [5–7], such as dementia [6, 8], schizophrenia (SCZ) [9, 10], major depressive disorder (MDD) [11], and also mortality [7, 12]. In addition, biological, environmental, and lifestyle factors associated with these disorders have been reported to correlate with BAG, for example, infections [13, 14], smoking [5], physical activity [15], and education level [16].

Genetic differences have been shown to explain a sizeable portion of interindividual variation in BAG. Twin-based heritability for BAG has been estimated to be as high as 0.66 [17], and single nucleotide polymorphism (SNP)-based heritability estimates are also relatively high—around 0.2 [6, 18]. Earlier gene-discovery efforts investigating genetic associations with BAG have found and examined two genomic loci in detail: one on chromosome 1 containing the potassium channel gene, *KCNK2*, and one in the chromosome 17 inversion region (17q21.31) [18, 19]. Genetic variants in these two regions together explain a negligible fraction of estimated SNP-heritability [18]. These results suggest that existing GWAS were potentially underpowered to fully characterize the genetic architecture, supported by studies using a conditional false discovery rate-based models yielding a larger set of associations [6]. Furthermore, Smith et al. [20] found a rich set of associations when investigating different facets of a multimodal brain age, suggesting that the interplay between genetic variants is complex.

Although BAG has been frequently associated with clinical conditions and health-related phenotypes and behaviors, the underlying genetic basis for the observed associations has seldom been investigated, possibly due to incomplete knowledge of the genetic architecture of the former. Furthermore, the causal relationships between BAG and brain disorders remain untapped. Mendelian randomization (MR) has become an attractive tool to interrogate cause-effect relationships between risk factors and disorders [21]. Two-sample MR models have been used to infer causal relations between hundreds of traits or diseases [22]. However, MR analyses targeting the causal relations

[1]Center for Lifespan Changes in Brain and Cognition (LCBC), Department of Psychology, University of Oslo, 0317 Oslo, Norway. [2]NORMENT, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, 0317 Oslo, Norway. [3]K.G. Jebsen Centre for Neurodevelopmental disorders, University of Oslo, Oslo, Norway. [4]Department of Psychology, University of Oslo, 0317 Oslo, Norway. [5]LREN, Centre for Research in Neurosciences, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne, 1015 Lausanne, Switzerland. [6]Department of Psychiatry, University of Oxford, OX1 2JD Oxford, UK. [7]Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, 72074 Tübingen, Germany. [8]Wellcome Centre for Integrative Neuroimaging (WIN FMRIB), University of Oxford, OX3 9DU Oxford, United Kingdom. ✉email: yunpeng.wang@psykologi.uio.no

Received: 5 January 2023 Revised: 18 April 2023 Accepted: 19 April 2023
Published online: 10 May 2023

between BAG and brain disorders and associated traits have been lacking [23].

In the present work, we improve the yield of genetic associations for BAG using three strategies: First, we estimate brain age using a state-of-the-art deep neural network architecture (SFCN-reg) trained on one of the largest samples assembled to date [5]. Then we perform a GWAS for BAG on out-of-sample predictions for a portion of the UK Biobank v3 data containing 28,104 unrelated individuals, about eight thousand more than earlier studies. Finally, we use two-sample MR to assess the genetic and causal relations between BAG and SCZ, bipolar disorder (BIP), Alzheimer's disease (AD), MDD, and Parkinson's disease (PD).

METHODS

Sample

All datasets used in the present study have been obtained from previously published studies that have been approved by their respective institutional review boards, research ethic committees, or other relevant ethic organizations.

We used UK Biobank imaging data (UKB, accession number 27412) released in 2019 in combination with a pre-compiled dataset from various sources (Supplementary Table S1) for brain age model training and estimation. For the downstream genetic analyses, we started with the initial 40,330 UKB participants that had undergone at least one brain scan (using baseline scans where more were available). We excluded those with recorded brain injury or neurological or psychiatric conditions, those failing standard image quality checks [5]. To quality check the genetic data, the protocol developed by the NealeLab (nealelab.is/uk-biobank) was strictly followed, in addition to participants who withdrew consent. After removing samples with failed image and genetics quality check, 28,104 unique participants remained.

Brain age estimation

A minimal preprocessing protocol was applied to all raw T1-weighted brain MRI images before brain age estimation [5]: The auto-recon pipeline from FreeSurfer 5.3 [24] was used to remove nonbrain tissue. The resulting volumes were reoriented to the standard FSL [25] orientation using fslreorient2std, and linearly registered to the 1 mm FSL (version 6.0) MNI152 template using FLIRT [26], with 6 degrees of freedom. For efficiency, during model fitting, we cropped a central cube spanning the voxels 6:173, 2:214, and 0:160 in the sagittal, coronal, and axial dimensions, respectively. Before modeling, all voxel intensities were normalized by a constant factor to produce values in the range [0, 1].

The data from all sources (Supplementary Table S1 and UKB) were split into five equally-sized and disjoint folds with comparable age ranges and sex distributions. Four of these folds were used for fitting the brain age model, and out-of-sample estimates were computed for the remaining fold. This procedure was repeated five times, resulting in out-of-sample brain age estimates for all participants. Next, BAG was calculated by subtracting chronological age from estimated brain age. The subsequent analyses were performed on the out-of-sample estimates of the UKB data (Supplementary Table S2).

Genome-wide association study

Imputed genotypes for the 28,104 participants were obtained from UKB (Category 100314, for further details see [27]). We excluded SNPs based on missing rate (>0.02), the Hardy-Weinberg Equilibrium test ($p < 10^{-6}$) and minor allele frequency (MAF; < 0.01). In total, ≈8.6 million SNPs were analyzed. Since we have observed apparent differences in predicted brain age across folds (Supplementary Fig. S1), a GWAS was performed on each hold-out fold separately using PLINK 1.90 beta [28]. The additive genetic model was assumed, and chronological age, sex and the top ten principal components were included as covariates, accounting for population structure. Association results for each hold-out fold of UKB along with distributions of BAG are shown in Supplementary Figure S1. These association results were then meta-analyzed using the inverse variance weighted model implemented in PLINK to identify SNPs that are associated with BAG. Supplementary Figure S2 shows the association QQ plot which indicated no noticeable genomic inflation.

Associated regions and genes

Association results were 'clumped' by the FUMA [29] web-service using the linkage disequilibrium (LD) structure from the 1000 Genomes projects phase 3 EUR dataset (1KGp3), with parameters *–clump-p 5e-8 –clump-2 1e-6 –clump-r2 0.1*. The standard 250 kilo-bases (kb) were used as the inter-region distance threshold. Genes whose genomic coordinates located within the boundaries of each region were assigned to the corresponding region. SNPs with the smallest association *p* values were taken as the lead SNPs for the corresponding regions. In addition, the gene that is closest to each lead SNP by genomic position was annotated using the Ensembl tool VEP [30] (Table 1).

Associated regions were fine-mapped using the FINEMAP [31] program. The LD structure from 1KGp3 was also used in this analysis. The default settings of FINEMAP were used, which compares causal models assuming one causal variant in each region to that assuming two, based on the estimated posterior probabilities (PP_1 versus PP_2). FINEMAP ranks all possible configurations in each model presented as 95% credible sets. The confidence of a variant belonging to a set was evaluated by posterior probabilities of inclusion (PPI). In the case of assuming one causal variant, each single variant was assigned a PPI. In the two causal variants cases, each pair of variants was assigned a PPI.

Post-GWAS annotations

Both FUMA and Garfield [32] were used for annotating associated SNPs. First, SNPs were assigned to genic elements (e.g., exon, intron, 3′ and 5′ untranslated regions, intergenic regions, etc.), and the enrichment of this assignment was tested by hypergeometric test (FUMA) or logistic regression models (Garfield). Expression levels of annotated genes to the associated SNPs were inspected in each of the 54 tissue types from the GTEx v8 dataset [33]. To further test if the identified variants affect expression levels of these genes the GTEx v8 eQTL portal (gtexportal.org) was searched. Data in this portal include the association statistics of SNPs with gene expressions in 49 different tissues. We took a conservative significant threshold to claim the existence of evidence as $p <= 0.05/49*8 = 1.3 \times 10^{-4}$. Moreover, detailed biological functions for proteins coded by these genes were manually searched in the NCBI Entrez Gene database [34] and the UniProtKB database [35].

Table 1. Genomic loci associated with BAG.

Locus	Lead SNP	POS	Gene	A1	A2	Beta	I^2	*P*
Chr3:183892867-183975709	rs73185796	183975709	CAMK2N2/ECE2	T	G	−0.29	0.0	2.53×10^{-8}
Chr4:38591172-38779512	rs13132853	38680015	KLF3	G	A	0.23	0.0	2.34×10^{-18}
Chr5:78388694-78451813	rs79107704	78388694	BHMT2	A	G	0.63	0.0	1.65×10^{-8}
Chr6:45407654-45511945	rs2790102	45432214	RUNX2	A	G	−0.15	0.0	8.92×10^{-9}
Chr8:124661974-124682971	rs7461069	124669029	KLHL38	A	G	−0.17	0.0	1.57×10^{-8}
Chr10:134544247-134597265	rs4880424	134584577	INPP5A	T	C	0.16	64.95	3.69×10^{-8}
Chr14: 88391116-88556525	rs17203398	88449847	GALC	C	G	−0.16	40.97	1.42×10^{-10}
Chr17: 43101281-44863413	rs2106786	43919096	MAPT	G	A	0.29	0.0	1.87×10^{-23}

Eight independent genomic loci significantly associated with brain age gap (BAG). Lead SNP rs-number, genomic position (in hg19 coordinates), effective allele (A1), the other allele (A2), effect size (Beta), meta-analysis heterogeneity (I^2), association strength (*P* value, *P*).

Genetic correlations between BAG and disorders

GWAS summary data for four disorders (SCZ [36], BIP [37], MDD [38], and AD [39]) were obtained from the Psychiatric Genomics Consortium (PGC, https://med.unc.edu/pgc/download-results). For each GWAS, the association results for European ancestral samples excluding samples from 23andMe were used (SCZ: n case=67,390, n control = 94,015; BIP: n case = 41,917, n control = 371,549; MDD: n case = 59,851, n control = 113,154; AD: n case = 71,880, n control = 383,378). The PD GWAS results were obtained from the fixed-effect meta-analysis performed by the International Parkinson Disease Genomics Consortium (IPDGC, n case = 33,674, n control = 449,056) [40].

Before post-GWAS analysis we processed the results from all GWAS using a standard protocol. Specifically, SNPs having a MAF < 0.05, or imputation INFO < 0.5, or ambiguous allelic coding (A/T, or C/G) were removed from subsequent analyses. The LD score model (ldsc) [41] was applied to estimate SNP-heritability and genetic correlations between BAG and disorders. Only high-quality SNPs published in the HapMap3 dataset were used for estimation. The LD score derived from the 1KGp3 was used as input to ldsc. The Benjamini-Hochberg False Discovery Rate (FDR) procedure was used to correct for multiple testing across disorders (FDR-corrected p < 0.05 was considered statistically significant).

To visualize polygenic enrichment, conditional QQ plots [42, 43] were made for BAG versus each disorder. In these plots, the QQ curves for the association statistics (-log10 *P* values) for BAG were stratified by the corresponding association strength for the conditioned disorder. As the association strength to the conditioned disorder increases, a successive leftward deflation in these curves indicates polygenic enrichment. Similarly, conditional QQ for each disorder versus BAG shows polygenic enrichment in the reverse direction.

Two-sample Mendelian randomization

To study the cause-effect relations between BAG and the five disorders, two sets of MR analyses were performed. The first set, using standard models, included the inverse-variance weighted model (IVW) [44], weighted median (wMed) [45], Egger regression (Egger) [46], and MR-PRESSO (PRESSO) [47]. For these analyses, only genome-wide significant SNPs ($p < 5\times10^{-8}$) to the exposure traits or disorders were used as potential instruments. The PLINK program and the LD structure of 1KGp3 dataset were used to select instruments with the following parameters, *--clump-kb 500 kb*, *--clump-p1 5×10^{-8}*, and *--clump-r2 0.01*. The TwoSampleMR package [19] was used for data harmonization and causal inference for the IVW, wMed, and Egger models. The same harmonized datasets were used as input to the MR-PRESSO software to assess outliers that may artificially affect MR estimates, i.e., SNPs that show horizontal pleiotropy to both BAG and disorders. Harmonized instrumental SNPs are shown in Supplementary Tables S6–S15.

The second set of models included the robust adjusted profile score (RAPS) [48] and the CAUSE models [49]. These models can make use of SNPs that show a suggestive level of association ($p < 10^{-3}$) with exposure to increase statistical power without incurring weak instrument bias in estimation. Although both models control for horizontal pleiotropy, CAUSE directly tests for a shared (correlated horizontal pleiotropy) versus a causal model for each relation [49]. The same instrument selection procedures used in the first set of models were used here, except that 10^{-3} was taken as the cut-off for selecting instruments, i.e., *–clump-p1 10^{-3}*.

As each of the six MR models has different assumptions that are difficult to verify in real data, a majority vote ensemble scheme was used to make conclusions for the existence of cause-effect relations: specifically, only when four or more models indicated a cause-effect relation (FDR adjusted $P < 0.05$) was such a relation considered causal.

In addition to applying multiple MR models, GWAS results for height measured for European samples [50] ($n = 253{,}288$; https://portals.broadinstitute.org/collaboration/giant/), for AD diagnosed in a Japanese sample [51] (n case= 3962 and n control = 4047) and an African sample [52] (n case = 2784 and n control=5 222) and for BIP diagnosed in a Japanese sample [53] (n case = 2964 and n control = 61,887) were used to corroborate MR findings. As commonly done in genetic studies, height was used as a negative exposure control to test if population stratification could generate spurious causal effects [54]. The non-European GWAS data were used to test if any observed causal effects generalize across ancestry groups, although with significantly smaller sample sizes.

RESULTS

We obtained accurate brain age estimates; mean absolute errors (MAEs) in each of the five disjoint folds were consistently below 2.5 years (Supplementary Table S2). This was consistent when we split the dataset into different subsets based on covariates (MAE = 2.40 in females compared to 2.53 in males; 2.40 in the youngest half compared to 2.52 in the oldest), although we observed a slight age bias (Supplementary Fig. S1). Based on the meta-analyzed GWAS results, we estimated a SNP heritability of 0.27 (standard error (SE) = 0.036) for BAG (Methods section). Our estimate is comparable to or higher than the two previously reported estimates (0.26, SE = 0.044 [6]; 0.19, SE = 0.02 [18]).

We identified eight independent genomic loci significantly associated with BAG (Fig. 1a, b and Table 1). Associations of lead SNPs in these regions to BAG are highly consistent in directions and effect sizes across the five folds (Supplementary Table S3). Among these loci, the one in the inversion region on chromosome 17 (lead SNP rs2106786), including the *MAPT* gene, has been previously reported [18, 19], although indexed by a different SNP (rs2435204). This SNP was also highly significant in our analysis ($p = 5.4 \times 10^{-21}$, beta = 0.27 years, effective-allele = G). The locus containing the *RUNX2* gene (lead by rs2790102: $p = 8.92 \times 10^{-9}$, beta = −0.15 years, effective-allele=A), which showed suggestive significance in Jonsson et al. [18], was genome-wide significant in the present study. The *RUNX2* gene codes for a master transcription factor which plays a critical role in skeletal development [55]. Among the remaining six novel loci, the rs79107704-A allele showed the largest association with BAG; one copy of this allele was associated with an average increase in brain age of 0.63 years (Table 1). This SNP is located 3405 bp downstream of the Betaine-homocystein S-methyltransferase 2 gene (*BHMT2*, Fig. 1b), a gene whose product is involved in choline metabolism during development [56]. Other protein-coding genes that are closest to the lead SNPs include those involved in calcium signaling (*CAMK2N2* and *INPP5A*) and metabolism and transcription regulation (*GALC*, *KLF3*, and *KLHL38*), both processes are implicated in biological ageing [57]. In Supplementary Table S5, we present detailed annotations of biological functions of each gene.

We further annotated these identified SNPs to nearby genes and regulatory elements (Methods). Most of the associated SNPs are in noncoding regions such as intergenic, intron or untranslated regions (Supplementary Figs. S3 and S4). Using the default parameters in FUMA [29], 54 unique genes were found to be implied by these significant associations by genomic position. The expression levels of these genes in the 54 tissue types from the GTEx v8 project [33] showed three remarkable patterns (Fig. 1c). The first set of genes expressed highly across almost all 54 tissues; the second set of genes showed low expression levels in most tissue types; and the last set, including eight genes, was highly expressed only in brain tissue (Fig. 1c), for example, *MAPT*, *GFAP*, and the Homeobox protein gene *NKX6-2*. These results suggest that BAG encodes coordinated physiological processes implicating both the brain and the peripheral systems.

To nominate causal variants in each locus we performed statistical fine-mapping [31] for regions around each lead SNP in Table 1 (Methods). Except for the locus on chromosome 14 which was not resolvable, all loci clearly indicated that the 95% credible sets suggest a causal model with one causal SNP, instead of two, i.e., the posterior probability for the 1-SNP set were larger than those of the 2-SNP sets (Supplementary Table S16). Furthermore, four credible sets indicated that the lead SNPs were also the causal ones (posterior inclusion probability (PPI_1) > 0.05 and >PPI of the second most probable SNP(PPI_2)) (Methods; Supplementary Table 16) but identifying the causal SNP for the remaining were difficult. For example, the *MAPT* locus on chromosome 17 and the *RUNX2* locus on chromosome 6 showed two SNPs having almost equal and small PPIs (i.e., <=0.05), indicating that the true causal

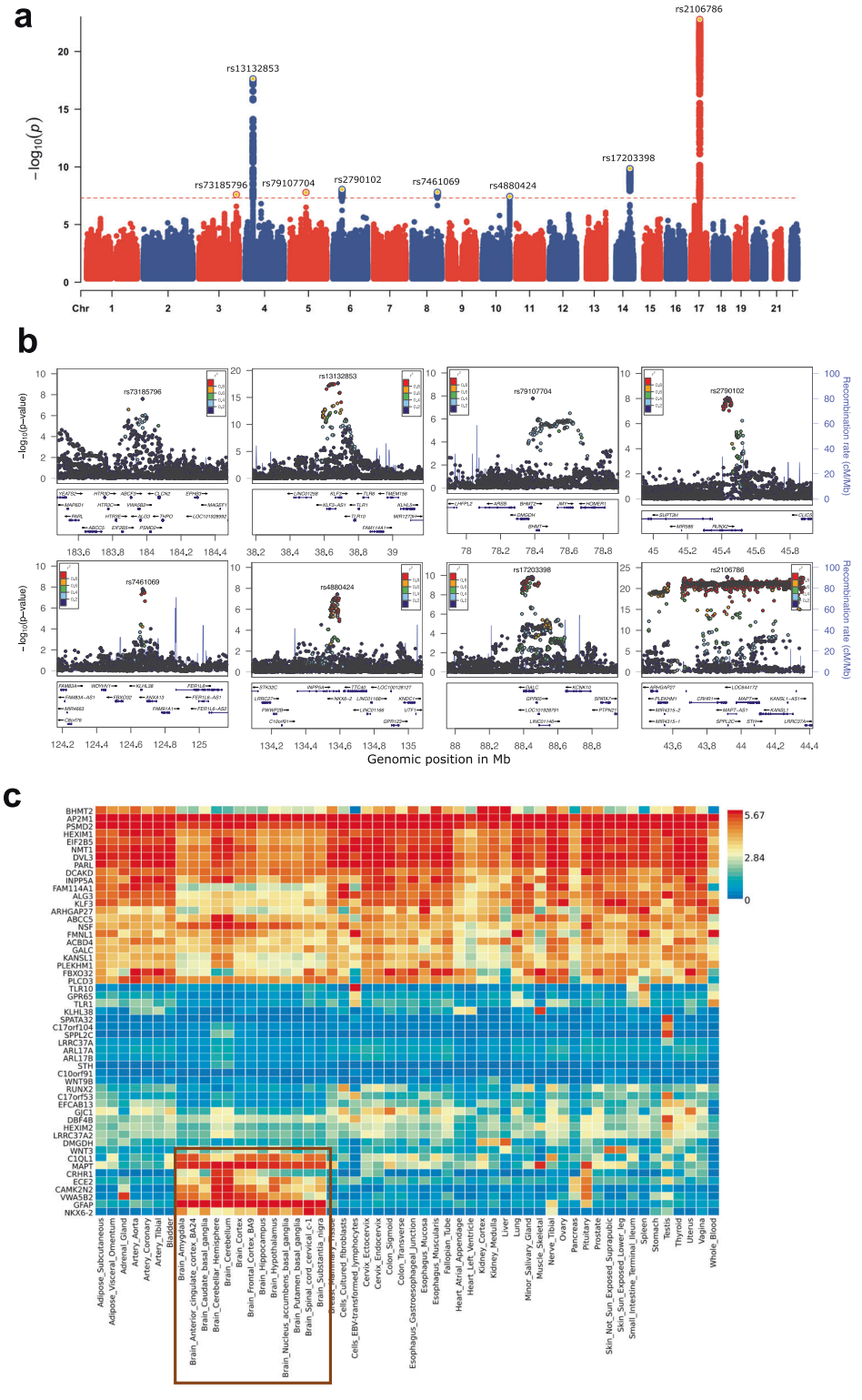


Fig. 1 Genetic associations for brain age gap. a The Manhattan plot of meta-analyzed association results for brain age gap (BAG). Chromosome numbers are shown on x axis, -log10 association p values on y axis and lead SNP rs-numbers in the plot. **b** Region plots for each of the eight associated regions. Genes located in each region are shown below each figure. Linkage disequilibrium r-squared values are indicated by colors; and recombination frequences by curves. **c** Expression levels of the annotated genes across tissues analyzed by the GTEx v8 study. Colors indicate average log2 transformed expression level in each tissue.

variants may be some untyped rare ones not investigated in this study. The clearest signal comes from the regions on chromosome 3 and 5, where the PPIs for the lead SNPs were much larger than that for the second most probable causal SNPs.

We investigate whether the statistically fine-mapped causal variants affect BAG through transcriptional regulatory mechanism using the GETx database (eQTL or sQTL data for 49 tissues; gtexportal.org; accession date 25 February 2023). Except rs79107704, the seven SNPs significantly affect the transcription levels of one or more nearby genes ($p < 1.3 \times 10^{-4}$; Supplementary Figs. S6–S12 and Tables S17–S22). Of note, rs2790102 and rs17203398 affect the only gene (RUNX2 and GALC, respectively). The other five affect the expression of two or more genes, particularly for rs2106786 that affect 37 unique genes or noncoding RNAs across all the 49 tissue types. In addition, this SNP also affects the splicing isoforms for 15 unique genes in the 49 tissue types (Supplementary Table S21-S22). This observed complex pattern makes it difficult to pin down the genes through which rs2106786 influences BAG.

We observed nominally significant genetic correlation between BAG and AD that did not survive FDR-correction ($r = 0.23$, SE = 0.1, $p = 0.02$, FDR adjusted p = 0.13) and no significant associations with any other of the four disorders (Fig. 2a). SNP heritability estimates for the five disorders were all significant but varied greatly; SCZ showed the largest (0.34, SE = 0.01) and AD showed the lowest (0.01, SE = 0.005) estimates. Bidirectional conditional QQ plots (Fig. 2b–d; Supplementary Figs. S13 and S14) showed that there was noticeable genetic enrichment for BIP conditional on BAG but not in the reverse direction. For AD and PD, both directions showed clear enrichment, surprisingly for PD that did not show significant genetic correlation with BAG ($r = -0.07$, $p = 0.42$).

We then performed extensive MR analyses using six different models to examine the existence of cause-effect relations between BAG and the five disorders (Methods). Figure 3a shows that BAG was only causally associated with PD, i.e., four out of the six MR models showed a negative relation with varying effect sizes (all with adjusted $p < 0.05$). One year increase in genetically predicted BAG was estimated to reduce the risk of PD by a log odds ratio from 1.4 (by Egger regression) to 0.02 (by MR-RAPS) (Supplementary Table S23). In the reverse direction (i.e., disorders as exposure), increased genetic risk for AD and BIP were causally associated with increased BAG (30 and 55 SNPs used as instruments, respectively); these estimated causal effects on BAG were relatively larger for AD than BIP (Fig. 3b; Supplementary Table S24).

A close investigation into the scatter plots of instrumental SNPs showed that the causal effect of AD on BAG was primarily driven by a SNP (rs59007384) in the APOE region, which was not identified as a horizontal pleiotropic instrument by MR-PRESSO (outlier test $p > 0.05$) (Fig. 3c); there were no extreme instruments identified for the BIP to BAG relation by MR-PRESSO (Fig. 3d) but Egger-regression indicated existence of horizontal pleiotropy (Egger intercept test: p = 0.017). The causal effects of BAG on PD were primarily driven by two SNPs in the inversion region on chromosome 17, effective alleles of these SNPs were associated with higher BAG and lower risk of PD (Fig. 3e). SNPs in the same region also drove the negative causal relation (not significant) from PD to BAG (Fig. 3f). However, both SNPs were flagged as horizontal pleiotropic instruments by MR-PRESSO ($p < 0.05$) and Egger-regression (Egger intercept test: $p = 0.03$ and 0.008, respectively). Therefore, the observed negative relations between BAG and PD are less likely to be causal.

We used the GWAS results for height of European samples and cross-ancestry MR analysis to corroborate the identified causal relations (Methods). We found no causal effect between BAG and height with any of the MR methods employed (all $p > 0.05$). Therefore, our observed AD and BIP to BAG relations are less likely to be driven by population stratification, i.e., both the exposure and outcome data originating from the same ancestry group. There was also no significant cross-ancestral causal effect detected using AD data from Japanese or African samples (IVW $p = 0.74$, 0,85, respectively), and BIP data from the Japanese sample to BAG (beta = 0.10, $p = 0.13$).

DISCUSSION

Combining the advantages of large samples and advanced models for brain age prediction, we confirmed that BAG is a heritable and polygenic trait, and estimated the genetic pleiotropy and causal genetic relations with major brain and mental disorders. We identified seven novel loci associated with BAG, in addition to confirming the previously reported *MAPT* loci [18, 19]. Although MR indicated that increased genetic risk for AD or BIP may be causally associated with higher BAG, our results demonstrate that individual variability and previously reported case-control differences in BAG only to a marginal degree should be attributed to the common genetic architecture previously associated with the respective diseases.

Functional annotation of the genes linked to the identified loci confirms that deviations in BAG are linked to complex processes encompassing multiple biological systems [20]. Although earlier work observed this variety when investigating different multimodal aspects of imaging data linked to brain age, our findings suggest it also exists when looking at a singular BAG computed from only T1-weighted MRI data. Our coarse division of the implied 54 genes into three groups indicates that only eight genes are specifically expressed in brain tissue. The remaining genes were either expressed in abundance across all tissue types tested, including the brain, or expressed at very low levels across all tissues. Nonetheless, the proteins coded by these nonbrain-specific genes have been implicated in brain-related disorders or traits (Supplementary Table S5). For example, among the genes we found to be expressed across all tissue types (group 1), mutations in *AP2M1* have been linked to epilepsy, intellectual developmental disorder, and seizures [58]; among the genes expressed in low levels across tissues (group 2), *STH* has been associated with frontotemporal dementia and 17q21.31 duplication syndrome [59, 60]. More importantly, we show that our fine-mapped causal SNPs affect the expression levels of these genes in multiple tissue types, providing testable molecular mechanisms for these genetic variants. In addition, although our analysis revealed no significant pathway enrichment, these 54 genes contribute to biological functions that include calcium signaling, protein metabolism, DNA damage repair, and general innate immune defense. Thus, our analyses highlight the role of these diverse sets of processes affecting the brain throughout life.

Prior work has shown higher BAG in patients with a multitude of disorders compared to healthy controls [5, 6, 8, 9, 11], and has documented partly overlapping genetic associations between BAG and clinical conditions [6]. However, the causal effects have remained unclear. Our MR approach suggested that genetically predicted risks for AD or BIP were causally associated with increased BAG. However, these relations were only weakly supported by genetic correlation analysis. One possible explanation for this weaker support from genome-wide signals (genetic correlation) in contrast to MR (significant associations only) might be due to heterogeneous genetic correlations across the genome, i.e., some genomic regions show positive correlations while others show negative correlations [61, 62]. In such a scenario, the net genetic correlations between the two traits are expected to be lower than regional correlations.

The causal effect of genetically predicted risk for AD on BAG was small but consistent in directions across the six MR models, four of which were significant after multiple-testing correction. For

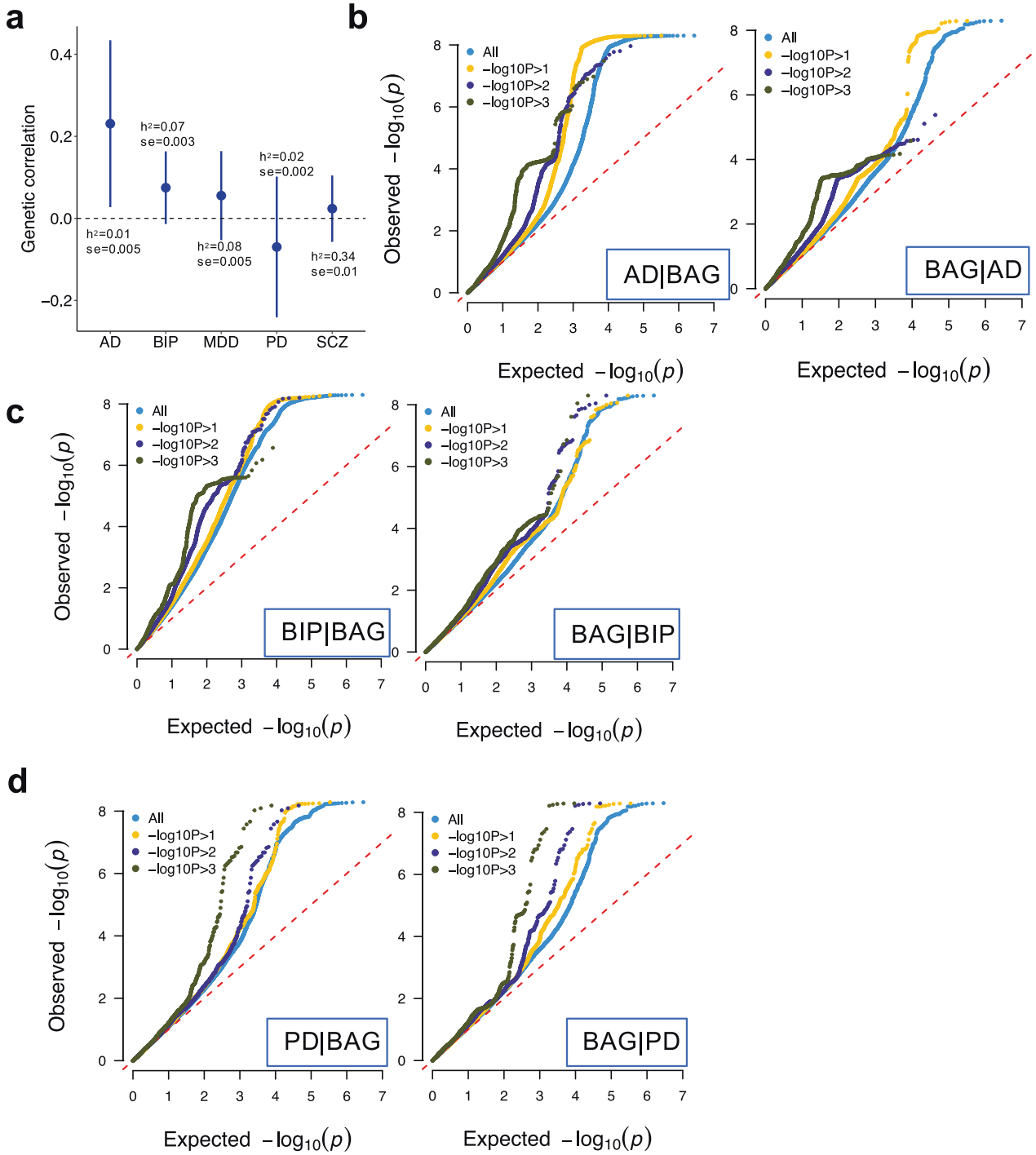


Fig. 2 Polygenic genetic overlap between brain age gap and disorders. a Genetic correlation between brain age gap and disorders computed by ldsc. SNP heritability and its standard error are indicated. **b**–**d** Conditional QQ plot between brain age gap and disorders in both directions. Colors are used to indicate different association strength to the conditioned traits, i.e., the ones indicated after the vertical bar in each figure. Dashed diagonal lines indicate expected null distributions. AD Alzheimer's disease, BIP bipolar disorder, MDD major depression disorder, PD Parkinson's disease, SCZ schizophrenia.

BIP, although four models showed significant effects, the CAUSE model suggested an opposite direction of effect to the other five models. Thus, we advise careful interpretation of this result. Our attempts of testing causal relations across ancestral groups led to largely null fundings for the AD to BAG relations. We believe these nonsignificant fundings are largely due to the lack of statistical power in the non-European GWAS [51–53].

The observed causal relations between genetically predicted risk of brain disorders and BAG are intriguing. One possible interpretation is that overt changes in the brain incurred by the disorders contribute to accelerated aging. Another possibility may be that lifestyle and health-related behaviors of patients with clinical conditions such as AD and BIP, *e.g.*, medication [63], may increase brain age. Yet another is that genetic variation associated

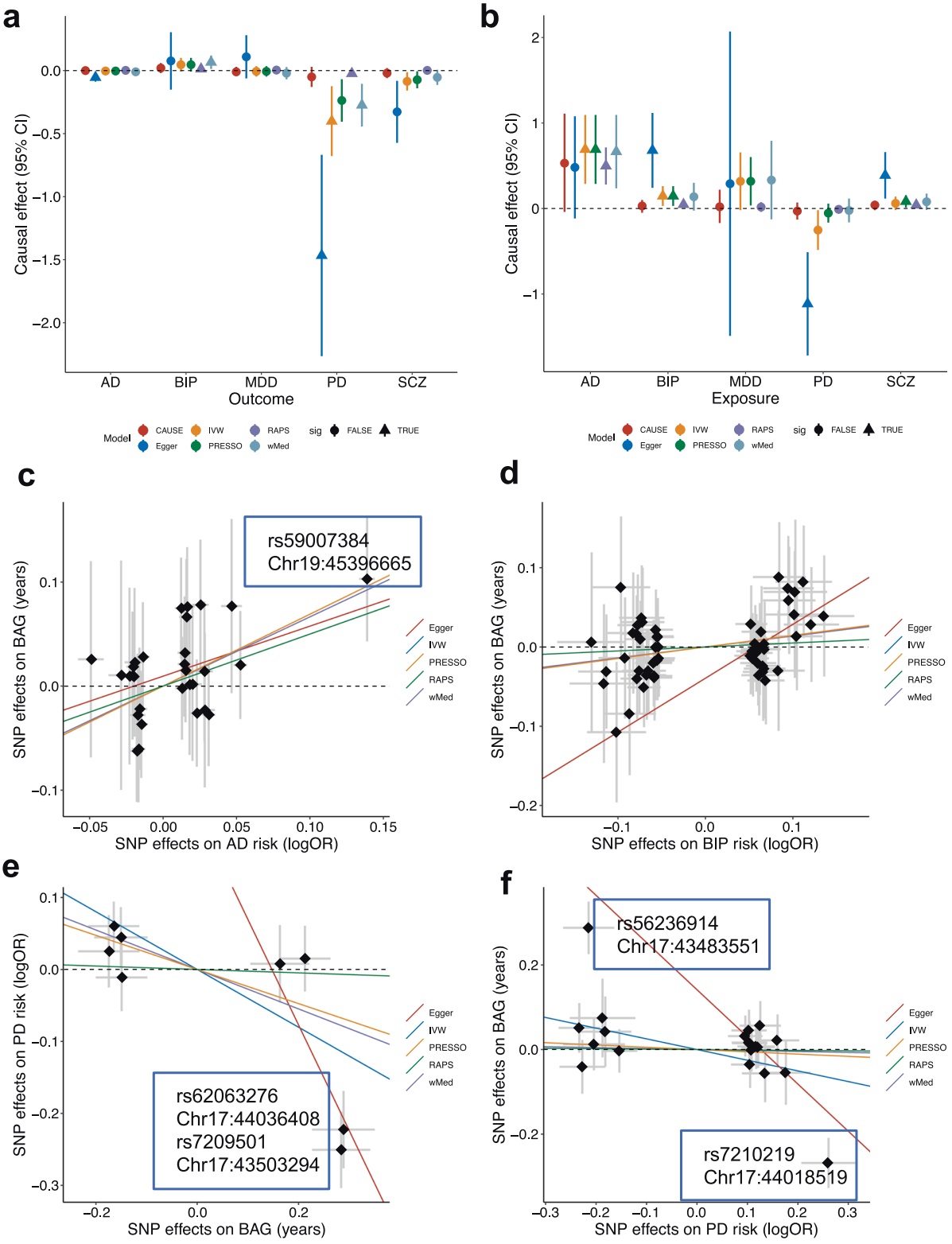


Fig. 3 Causal inference between brain age gap and disorders. a Causal effect of brain age gap (BAG) on risk of disorders; **b** Causal effect of genetic risk of disorders on BAG. Colors indicate different models; triangle indicates significant effect after false discovery correction. Estimated standard errors for each effect are aslo shown. **c** Scatter plots of SNP effects on AD (x axis) and BAG (y axis). **d** Scatter plots of SNP effects on BIP (x axis) and BAG (y axis). **e** Scatter plots of SNP effects BAG (x axis) and PD (y axis). **f** Scatter plots of SNP effects on PD (x axis) and BAG (y axis). Causal effects estimated by the five models (except CAUSE) are shown by fitted lines; slopes of these lines indicate causal effect sizes. Exceptional SNPs are marked by boxes that include SNP rs-numbers and genome location in the hg19 coordinates. AD Alzheimer's disease, BIP bipolar disorder, MDD major depression disorder, PD Parkinson's disease, SCZ schizophrenia.

with clinical traits may influence the brain early in life. Given the comparable sample sizes to the GWAS of AD and BIP and the widely observed clinical correlations, surprisingly, no genetic nor causal relations of SCZ and MDD with BAG were found. On the one hand, this may suggest that previously reported case-control differences in BAG do not reflect causal relations, but rather a combination of indirect and confounding factors. For example, smoking and physical exercise have been associated both with MDD and SCZ [64–67] and brain age [5, 66]. Alternatively, it has been shown that both BAG [20] and psychiatric disorders are highly heterogeneous phenotypes [68, 69], and thus further identification and characterization of the causal relations may require even larger, and carefully screened, samples. It is also worth noting that while the sample sizes for the disorders are large, our BAG GWAS sample is relatively small. Thus, our null findings in the direction from BAG to disorders may be due to too weak instruments [70].

Our initial results showed weak evidence of a causal relation between BAG and PD, corroborating two recent studies which reported a weak correlation between BAG and PD [71, 72]. Striking patterns of enrichment between the two were shown in the conditional QQ plots and four out of six MR models indicated that genetically predicted BAG may have protective effect from PD. However, we found that these relations were completely caused by the *MAPT* gene region on chromosome 17: After removing chromosome 17 from our analyses, no enrichment was observed in either direction (Supplementary Fig. S15). In addition, instrumental SNPs in this region were detected by MR-PRESSO as horizontal pleiotropic SNPs, i.e., affecting BAG and PD through independent biological pathways. Reperforming MR analysis excluding these outlier SNPs confirmed null causal relations. Thus, we conclude that we found no evidence for causal relations between genetically predicted risk for PD and BAG. Our analytic procedures also highlight the importance of triangulation and converging evidence in causal inference analysis [73].

While the present study advances current knowledge regarding the genetic architecture of and causal contributions to BAG, the results should be interpreted with caution. Although we confirm previously reported genetic associations with BAG, e.g., the *MAPT* gene locus [18, 19], our sample overlaps with previous ones—which were also based on UK Biobank data. We attempted replicating our findings in three independent but small samples (n ranges from 321 to 702; Supplementary Analysis and Table S4) but no clear replications were achieved. Therefore, independent large-scale samples are needed for replication. We used a simple voting schema across six different MR models to infer causal relations between genetically predicted BAG and brain disorders. Furthermore, as only eight independent loci showed significant associations with BAG, other models [74] that require large number of genome-wide significant instruments were considered not applicable. However, it should be noted our simple voting approach may not be the most efficient strategy for identifying causal effects. Formal development of ensemble methods, such as bagging [75], may provide better grounds for precise interpretation. Furthermore, our BAG GWAS is still smaller than GWAS performed for the disorders, which may partly explain the lack of causal effects of BAG on brain disorders. Another limitation is that we were unable to obtain independent data to perform three sample MR analysis, a model that can account for the winner's curse bias in two sample MR models. Therefore, to increase our confidence in the identified relations, large-scale data for BAG, and replications in independent datasets are needed. Relatedly, our estimation of brain age was based on cross-sectional samples, which makes its interpretation nontrivial [76], and studies built on longitudinal data could help disentangle its complexities. Finally, although we refer to our brain age estimation in general terms, it is based on T1-weighted MRI data only. The brain is a complex and heterogeneous organ, and different imaging modalities are known to capture different aspects of the naturally occurring variation. Thus, studies relying on other modalities, either independently or in combination, could reveal a broader set of associations [77].

In conclusion, the present study increases the yield of genetic associations with brain age to eight genomic loci; implicated genes indicate involvement of calcium signaling, DNA damage repair, protein metabolism, and general innate immune defense. Our analysis did not provide evidence of a causal relationship between BAG and the included clinical conditions, and their interactions remain unclear.

REFERENCES

1. Franke K, Gaser C. Ten years of BrainAGE as a neuroimaging biomarker of brain aging: what insights have we gained? Front Neurol. 2019;10:789.
2. Cole JH, Franke K. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends Neurosci. 2017;40:681–90.
3. Franke K, Luders E, May A, Wilke M, Gaser C. Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. NeuroImage. 2012;63:1305–12.
4. Smith SM, Vidaurre D, Alfaro-Almagro F, Nichols TE, Miller KL. Estimation of brain age delta from brain imaging. NeuroImage. 2019;200:528–39.
5. Leonardsen EH, Peng H, Kaufmann T, Agartz I, Andreassen OA, Celius EG, et al. Deep neural networks learn general and clinically relevant representations of the ageing brain. NeuroImage. 2022;256:119210.
6. Kaufmann T, van der Meer D, Doan NT, Schwarz E, Lund MJ, Agartz I, et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. Nat Neurosci. 2019;22:1617–23.
7. Cole JH, Ritchie SJ, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle N, et al. Brain age predicts mortality. Mol Psychiatry. 2018;23:1385–92.
8. Gaser C, Franke K, Klöppel S, Koutsouleris N, Sauer H, Alzheimer's Disease Neuroimaging I. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. PloS one. 2013;8:e67346.
9. Schnack HG, van Haren NEM, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. Am J Psychiatry. 2016;173:607–16.
10. Constantinides C, Han LK, Alloza C, Antonucci L, Arango C, Ayesa-Arriola R, et al. Brain ageing in schizophrenia: evidence from 26 international cohorts via the ENIGMA Schizophrenia consortium. Mol Psychiatry. 2023;28:1201–9.
11. Han LKM, Dinga R, Hahn T, Ching CRK, Eyler LT, Aftanas L, et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. Mol Psychiatry. 2021;26:5124–39.
12. Elliott ML, Belsky DW, Knodt AR, Ireland D, Melzer TR, Poulton R, et al. Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. Mol Psychiatry. 2021;26:3829–38.
13. Kuhn T, Kaufmann T, Doan NT, Westlye LT, Jones J, Nunez RA, et al. An augmented aging process in brain white matter in HIV. Hum Brain Mapp. 2018;39:2532–40.
14. Cole JH, Underwood J, Caan MWA, De Francesco D, van Zoest RA, Leech R, et al. Increased brain-predicted aging in treated HIV disease. Neurology. 2017;88:1349–57.
15. Steffener J, Habeck C, O'Shea D, Razlighi Q, Bherer L, Stern Y. Differences between chronological and brain age are related to education and self-reported physical activity. Neurobiol Aging. 2016;40:138–44.
16. Wrigglesworth J, Ward P, Harding IH, Nilaweera D, Wu Z, Woods RL, et al. Factors associated with brain ageing - a systematic review. BMC Neurol. 2021;21:312.
17. Cole JH, Poudel RPK, Tsagkrasoulis D, Caan MWA, Steves C, Spector TD, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. NeuroImage. 2017;163:115–24.
18. Jonsson BA, Bjornsdottir G, Thorgeirsson TE, Ellingsen LM, Walters GB, Gudbjartsson DF, et al. Brain age prediction using deep learning uncovers associated sequence variants. Nat Commun. 2019;10:5409.
19. Ning K, Zhao L, Matloff W, Sun F, Toga AW. Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants. Sci Rep. 2020;10:10.
20. Smith SM, Elliott LT, Alfaro-Almagro F, McCarthy P, Nichols TE, Douaud G, et al. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. eLife. 2020;9:e52677.
21. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32:1–22.
22. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.
23. Kolbeinsson A, Filippi S, Panagakis Y, Matthews PM, Elliott P, Dehghan A, et al. Accelerated MRI-predicted brain ageing and its associations with cardiometabolic and brain disorders. Sci Rep. 2020;10:19940.

24. Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, et al. A hybrid approach to the skull stripping problem in MRI. NeuroImage. 2004;22:1060–75.
25. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage. 2012;62:782–90.
26. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001;5:143–56.
27. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9.
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
29. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
30. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:122.
31. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32:1493–501.
32. Yap CX, Henders AK, Alvares GA, Wood DLA, Krause L, Tyson GW, et al. Autism-related dietary preferences mediate autism-gut microbiome associations. Cell. 2021;184:5916–31.e5917.
33. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
34. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2011;39:D52–D57. suppl_1
35. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A UniProtKB/Swiss-Prot. In: Edwards D (ed). *Plant Bioinformatics: Methods and Protocols*. Humana Press: Totowa, NJ, 2007, pp 89–112.
36. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022;604:502–8.
37. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021;53:817–29.
38. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50:668–81.
39. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019;51:404–13.
40. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18:1091–102.
41. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
42. Lo MT, Wang Y, Kauppi K, Sanyal N, Fan CC, Smeland OB, et al. Modeling prior information of common genetic variants improves gene discovery for neuroticism. Hum Mol Genet. 2017;26:4530–9.
43. Chen CH, Wang Y, Lo MT, Schork A, Fan CC, Holland D, et al. Leveraging genome characteristics to improve gene discovery for putamen subcortical brain structure. Sci Rep. 2017;7:15736.
44. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012;41:161–76.
45. Bowden J, Davey, Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.
46. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
47. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693–8.
48. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Ann Stat. 2020;48:1742–69.
49. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52:740–7.
50. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46:1173–86.
51. Shigemizu D, Mitsumori R, Akiyama S, Miyashita A, Morizono T, Higaki S, et al. Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. Transl psychiatry. 2021;11:151.
52. Kunkle BW, Schmidt M, Klein H-U, Naj AC, Hamilton-Nelson KL, Larson EB, et al. Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. JAMA Neurol. 2021;78:102–13.
53. Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N, et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. Mol Psychiatry. 2018;23:639–47.
54. Sanderson E, Richardson TG, Hemani G, Davey, Smith G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. Int J Epidemiol. 2021;50:1350–61.
55. Lam K, Zhang DE. RUNX/CBF Transcription Factors☆. *Reference Module in Biomedical Sciences*. Elsevier, 2015.
56. Kotb M, Geller AM. Methionine adenosyltransferase: Structure and function. Pharmacol Therap. 1993;59:125–43.
57. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153:1194–217.
58. Helbig I, Lopez-Hernandez T, Shor O, Galer P, Ganesan S, Pendziwiat M, et al. A recurrent missense variant in AP2M1 impairs Clathrin-mediated endocytosis and causes developmental and epileptic encephalopathy. Am J Hum Genet. 2019;104:1060–72.
59. Mc Cormack A, Taylor J, Te Weehi L, Love DR, George AM. A case of 17q21.31 microduplication and 7q31.33 microdeletion, associated with developmental delay, microcephaly, and mild dysmorphic features. Case Rep. Genet. 2014;2014:658570.
60. Arbogast T, Iacono G, Chevalier C, Afinowi NO, Houbaert X, van Eede MC, et al. Mouse models of 17q21.31 microdeletion and microduplication syndromes highlight the importance of Kansl1 for cognition. PLOS Genet. 2017;13:e1006886.
61. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat Commun. 2019;10:2417.
62. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. Nat Rev Genet. 2019;20:567–81.
63. Van Gestel H, Franke K, Petite J, Slaney C, Garnham J, Helmick C, et al. Brain age in bipolar disorders: Effects of lithium treatment. Aust NZ J Psychiatry. 2019;53:1179–88.
64. Pasco JA, Williams LJ, Jacka FN, Ng F, Henry MJ, Nicholson GC, et al. Tobacco smoking as a risk factor for major depressive disorder: population-based study. Br J Psychiatry. 2008;193:322–6.
65. Winterer G. Why do patients with schizophrenia smoke? Curr Opinion Psych 2010;23:112–9.
66. Sanders A-M, Richard G, Kolskår K, Ulrichsen KM, Kaufmann T, Alnæs D, et al. Linking objective measures of physical activity and capability with brain structure in healthy community dwelling older adults. NeuroImage: Clin. 2021;31:102767.
67. Brokmeier LL, Firth J, Vancampfort D, Smith L, Deenik J, Rosenbaum S, et al. Does physical activity reduce the risk of psychosis? A systematic review and meta-analysis of prospective studies. Psychiatry Res. 2020;284:112675.
68. Wolfers T, Doan NT, Kaufmann T, Alnaes D, Moberget T, Agartz I, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. JAMA Psychiatry. 2018;75:1146–55.
69. Wolfers T, Rokicki J, Alnæs D, Berthet P, Agartz I, Kia SM, et al. Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. Hum Brain Mapp. 2021;42:2546–55.
70. Taschler B, Smith SM, Nichols TE. Causal inference on neuroimaging data with Mendelian randomisation. NeuroImage. 2022;258:119385.
71. Eickhoff CR, Hoffstaedter F, Caspers J, Reetz K, Mathys C, Dogan I, et al. Advanced brain ageing in Parkinson's disease is related to disease duration and individual impairment. Brain Commun. 2021;3:fcab191.
72. Charissé D, Erus G, Pomponio R, Gorges M, Schmidt N, Schneider C, et al. Brain age and Alzheimer's-like atrophy are domain-specific predictors of cognitive impairment in Parkinson's disease. Neurobiol aging. 2022;109:31–42.
73. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45:1866–86.
74. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nat Rev Methods Prim. 2022;2:6.
75. Büchlmann P, Yu B. Analyzing bagging. Ann Stat. 2002;30:927–61.
76. Vidal-Pineiro D, Wang Y, Krogsrud SK, Amlien IK, Baaré WFC, Bartres-Faz D, et al. Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change. eLife. 2021;10:e69995.
77. Rokicki J, Wolfers T, Nordhøy W, Tesli N, Quintana DS, Alnæs D, et al. Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. Hum Brain Mapp. 2021;42:1714–26.

ACKNOWLEDGEMENTS

We thank for the computational resources provided by UNINETT Sigma2-the National Infrastructure for High-Performance Computing and Data Storage in Norway – with project no. (nn9769k/ns9769k), the PGC consortium and the International Parkinson Disease Genomics Consortium (IPDGC) for sharing of their GWAS results.

FUNDING

This study is supported by Norwegian Research Council (No. 223273, 298646, 300767, 302854), the UiO:Life Science Convergence environment, University of Oslo, Norway (4MENT), the South-Eastern Norway Regional Health Authority (2019101), KG Jebsen Stiftelsen, and the European Research Council under the European Union's Horizon 2020 research and Innovation program (ERC StG, Grant 802998).

AUTHOR CONTRIBUTIONS

EL, TW, and YW designed this study. EL, DVP, OF, AAS, OI, AMG, and YW performed data analysis. EL, DVP, JMR, OI, TK, LTW, OOA, TW, BT, SMS, and YW interpreted the results. EL, DVP, JMR, OI, and YW prepared the first draft of the manuscript. All authors contributed and approved the final draft.

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/s41380-023-02087-y.

Correspondence and requests for materials should be addressed to Yunpeng Wang.

Reprints and permission information is available at http://www.nature.com/reprints

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

Supporting Information
Genetic architecture of brain age and its casual relations with brain disorders
Leonardsen et al.

Table of Contents

Figure S1. Genetic association to brain age gap in each fold. **4**

Figure S2. QQ plot for the genetic association to brain age gap. **5**

Figure S3. Genic annotations for associated SNPs. **6**

Figure S4. Annotations for associated SNPs. **7**

Figure S5. Gene based association results. **8**

Figure S6. eQTL annotations for rs73185796. ..... **9**

Figure S7. eQTL annotations for rs13132853. ..... **10**

Figure S8. eQTL annotations for rs2790102. ..... **11**

Figure S9. eQTL and sQTL annotations for rs7461069. ..... **12**

Figure S10. eQTL and sQTL annotations for rs4880424. ..... **13**

Figure S11. eQTL and sQTL annotations for rs17203398. ..... **14**

Figure S12. eQTL annotations for rs2106786. ..... **15**

Figure S13. Conditional QQ plot for MDD vs BAG. **16**

Figure S14. Conditional QQ plot for SCZ vs BAG. **17**

Figure S15. Conditional QQ plot for PD vs BAG excluding chr17. ..... **18**

Table S1. Non-UK Biobank model training datasets. **19**

Table S2. Characteristics for the five folds data from UK Biobank. **20**

Table S3. Statistics for lead SNPs of each associated region in each fold. **21**

Table S4. Statistics for lead SNPs of each associated region in each replication samples. **22**

Table S5 Function annotation for genes in Figure 1c. **23**

Table S6. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to SCZ MR. **31**

Table S7. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to BIP MR. **32**

Table S8. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to MDD MR. **33**

Table S9. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to AD MR. **34**

Table S10. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to PD MR. **35**

Table S11. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the SCZ to BAG MR. **36**

Table S12. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BIP to BAG MR. **45**

Table S13. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the MDD to BAG MR. **48**

Table S14. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the AD to BAG MR. **49**
Table S15. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the PD to BAG MR. **51**
Table S16. Statistical finemapping for associated loci. **52**
Table S17. eQTL statistics for rs73185796 from GTEx v8 portal. **53**
Table S18. eQTL statistics for rs13132853 from GTEx v8 portal. **54**
Table S19. eQTL statistics for rs2790102 from GTEx v8 portal. **55**
Table S20. eQTL statistics for rs17203398 from GTEx v8 portal. **56**
Table S21. eQTL statistics for rs2106786 from GTEx v8 portal. **57**
Table S22. sQTL statistics for rs2106786 from GTEx v8 portal. **75**
Table S23. MR results using BAG as exposure. **82**
Table S24. MR results using disorders as exposure. **83**
References **84**

Replication analysis

Genetic data

We attempted to replicate the eight associations to BAG using three independent datasets, the Alzheimer's Disease Neuroimaging Initiative phase 1 (ADNI I) and 2 (ADNI II) and a healthy control dataset (Local sample) that have been previously published in Ripke et al[1] . Out local sample has been imputed by psychiatry genomic consortium using their standardized pipeline Ricopili[2]. However, for the two ADNI datasets, only genotyped data are publicly available. We imputed these genotypes to the 1000 Genomes Projects reference data, using Genipe[3]. The default parameters of Genipe were used. After imputation, we excluded SNPs having MAF < 0.01, HWE p <10-6, and imputation R-sq <0.3 from subsequent analysis.

Image data

MRI data for ANDI sample were obtained from http://adni.loni.usc.edu/ with permission. Brain scans for our local sample were performed at Oslo University Hospital and published in previous works[4-6]. The same protocols of image preprocessing and BAG estimation used in discovery sample were applied to these replication sample.

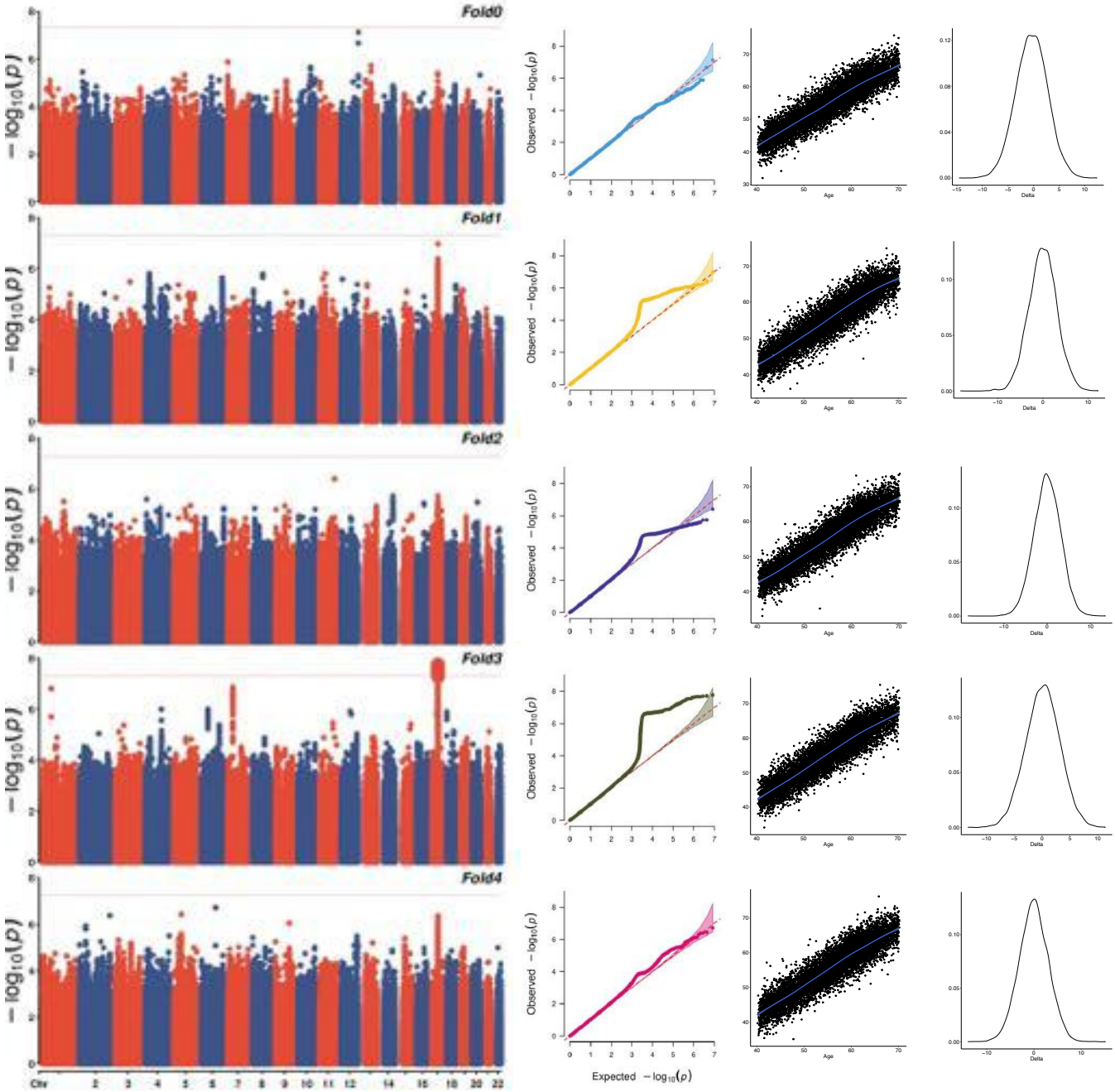
Replication association studies

We performed association test for each of the eight SNPs in each of the three replication datasets separately. For each analysis, chronological age, sex and the top ten principal components were included as covariates. Association results are shown in Supplementary Table S4.

Supplementary Acknowledgement

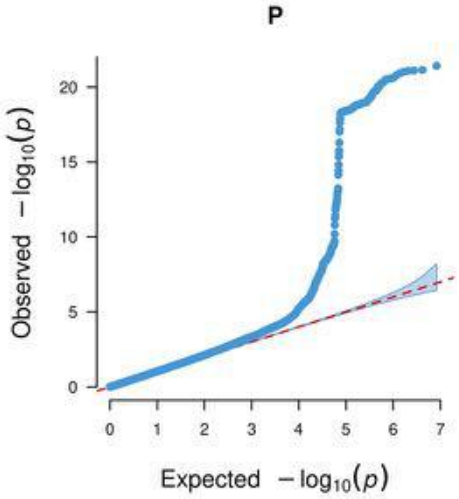
We thank the ADNI consortium for making their data publicly accessible. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. ADNI consists of 4 waves, the later is still ongoing (ADNI 3). A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Data collection and sharing for this project was funded by the ADNI (NIH Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada and the Saguenay Yout Study. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Figure S1. Genetic association to brain age gap in each fold.



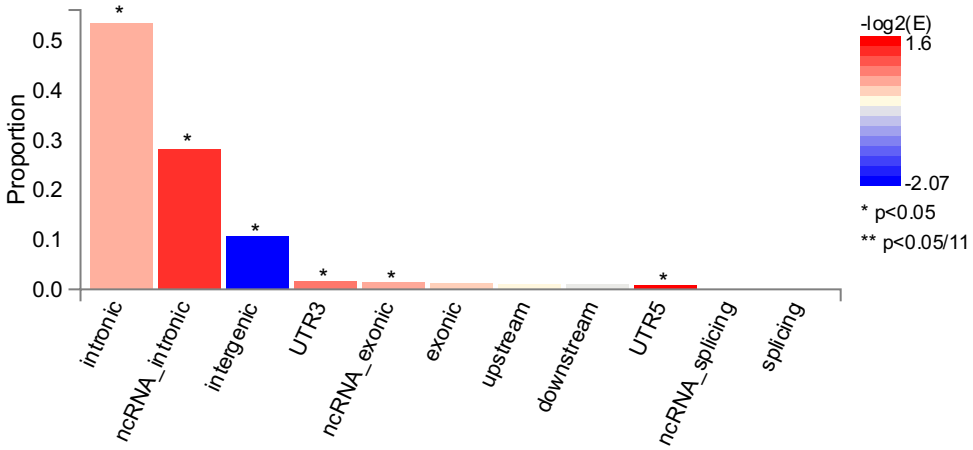
From left to right, GWAS association Manhattan plot and QQ plot, the scatter plot for estimated brain age (y axis) vs. chronological age (x axis), distribution for brain age gap (delta) for each fold of the UKB data (each row). In Manhattan plots, chromosome numbers are on x axes and -log10 association p value on the y axes. In QQ plots, the expected and observed p value on the -log10 scale are shown on the x and y axes. Blue lines in the scatter plots indicated a fitted loess model for each fold.

Figure S2. QQ plot for the genetic association to brain age gap.



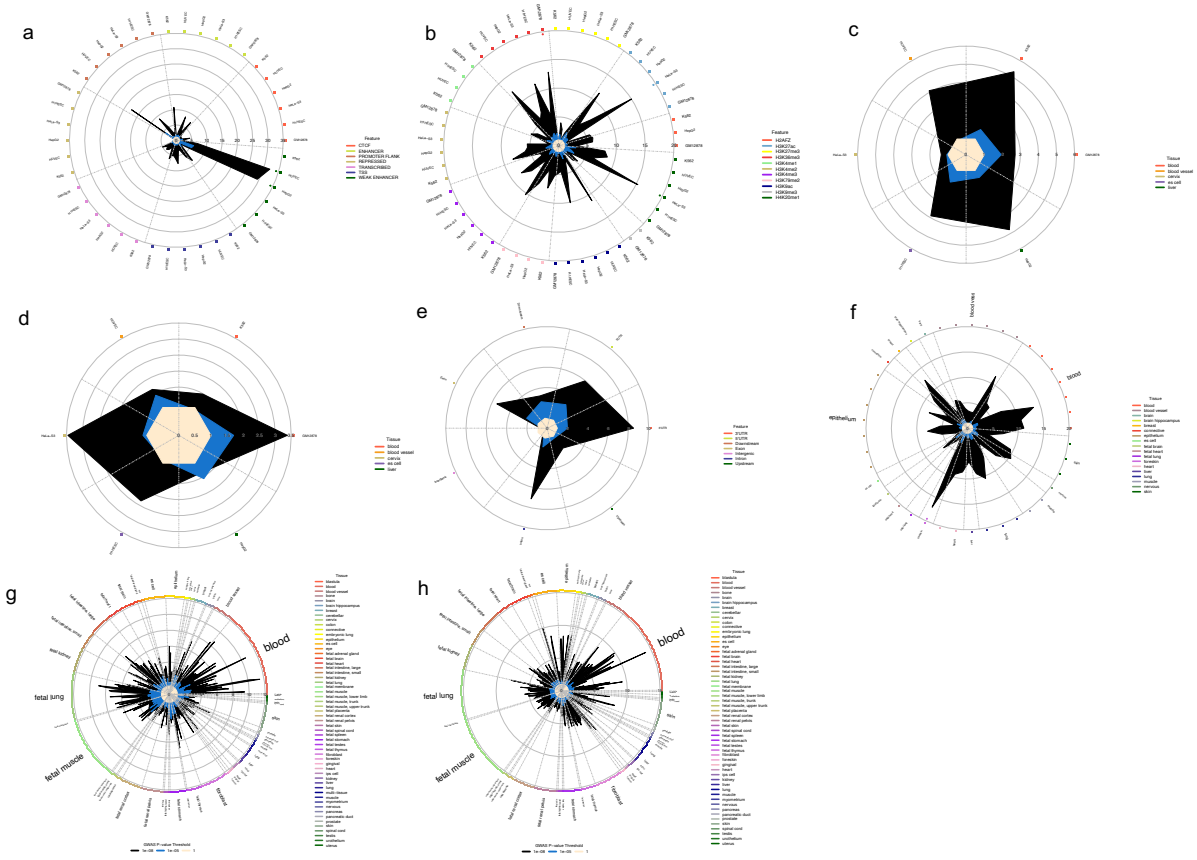
QQ plot for the results of meta-analysis over the GWAS for BAG in each of the five folds of UKB data.

Figure S3. Genic annotations for associated SNPs.



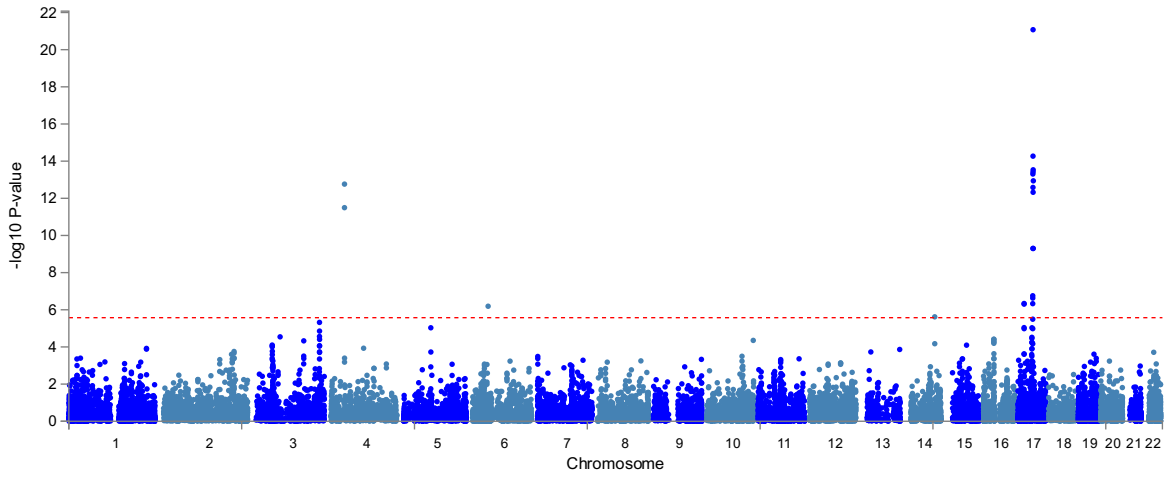
Annotations of associated SNPs to genic elements by FUMA[7]. UTR3, 3 prime untranslated region; ncRNA, non-coding RNA; UTR5, 5 prime untranslated region. Enrichment p values are based on hypergeometric test.

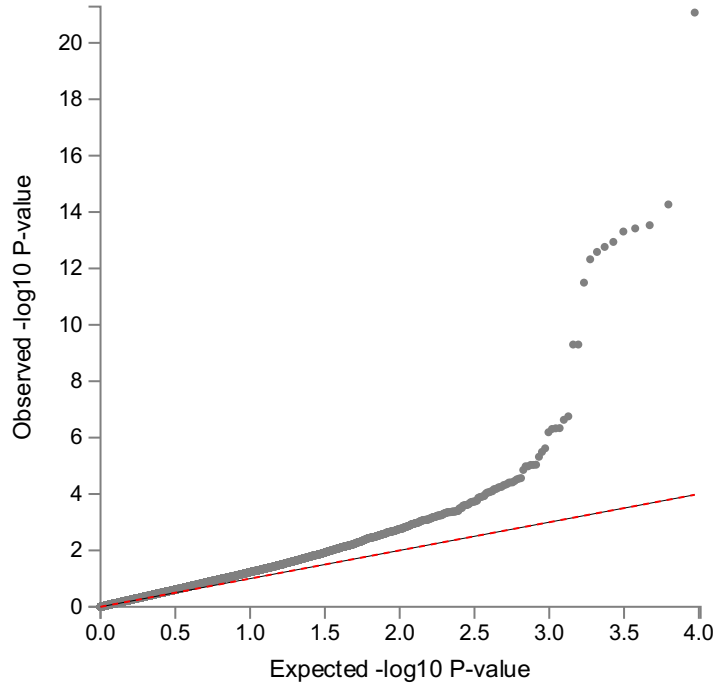
Figure S4. Annotations for associated SNPs.



Annotations of associated SNPs to genic elements by Garfield[8].

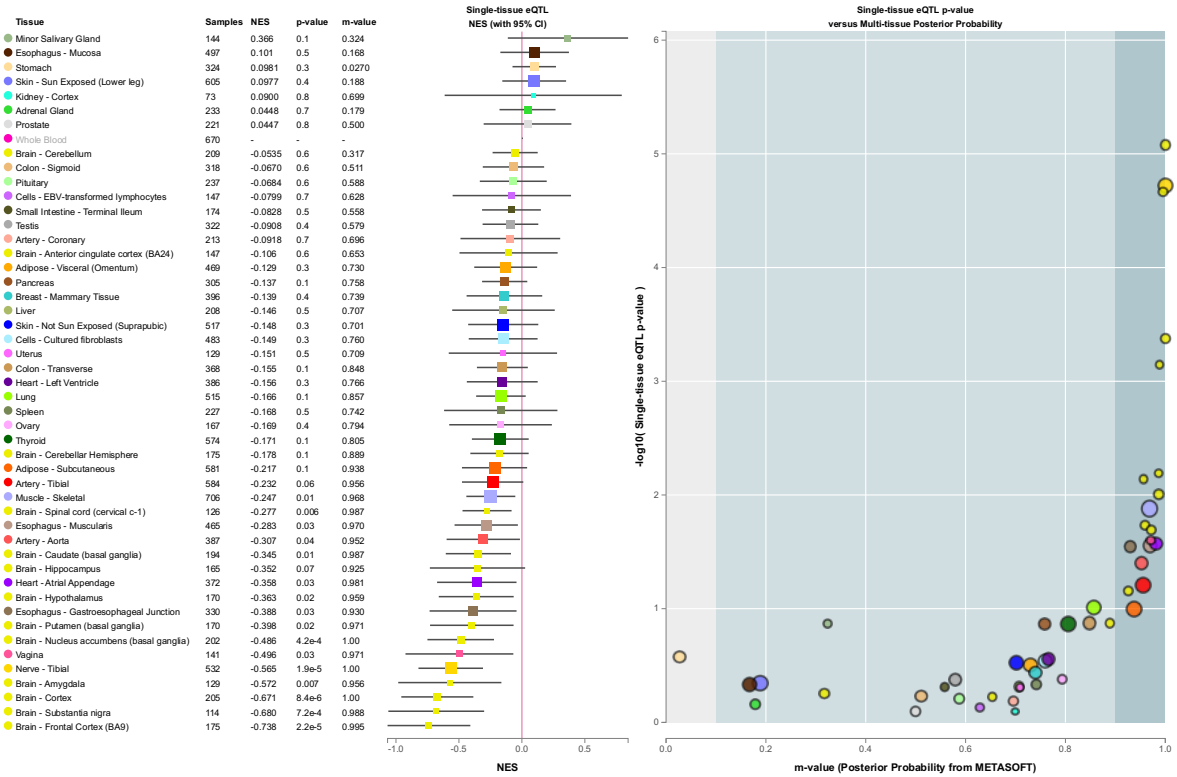
Figure S5. Gene based association results.





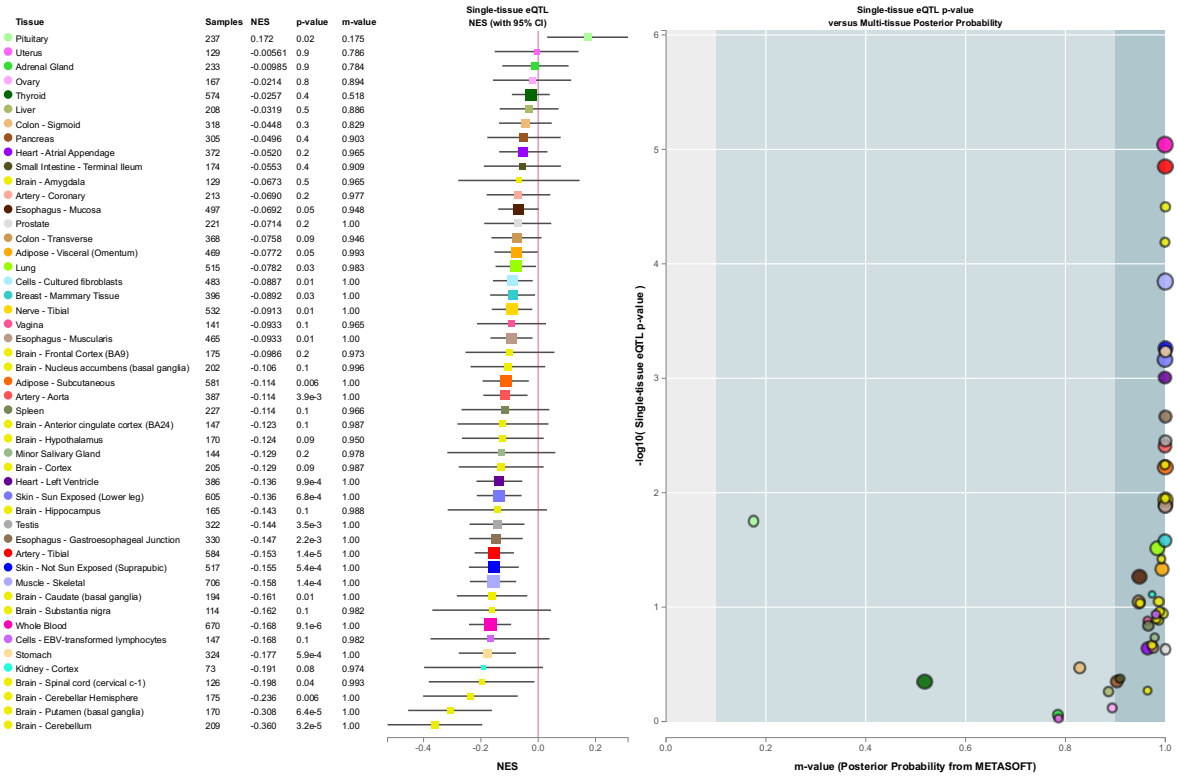
Manhattan plot and QQ plot for gene-based association statistics for BAG. Dashed line in the Manhattan plot indicates the genome wide significant threshold, and, in the QQ plot indicates the null association.

Figure S6. eQTL annotations for rs73185796.



The SNP rs73185796 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *VWA5B2*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL. Supplementary Tables XXX include numeric values for other eGenes.

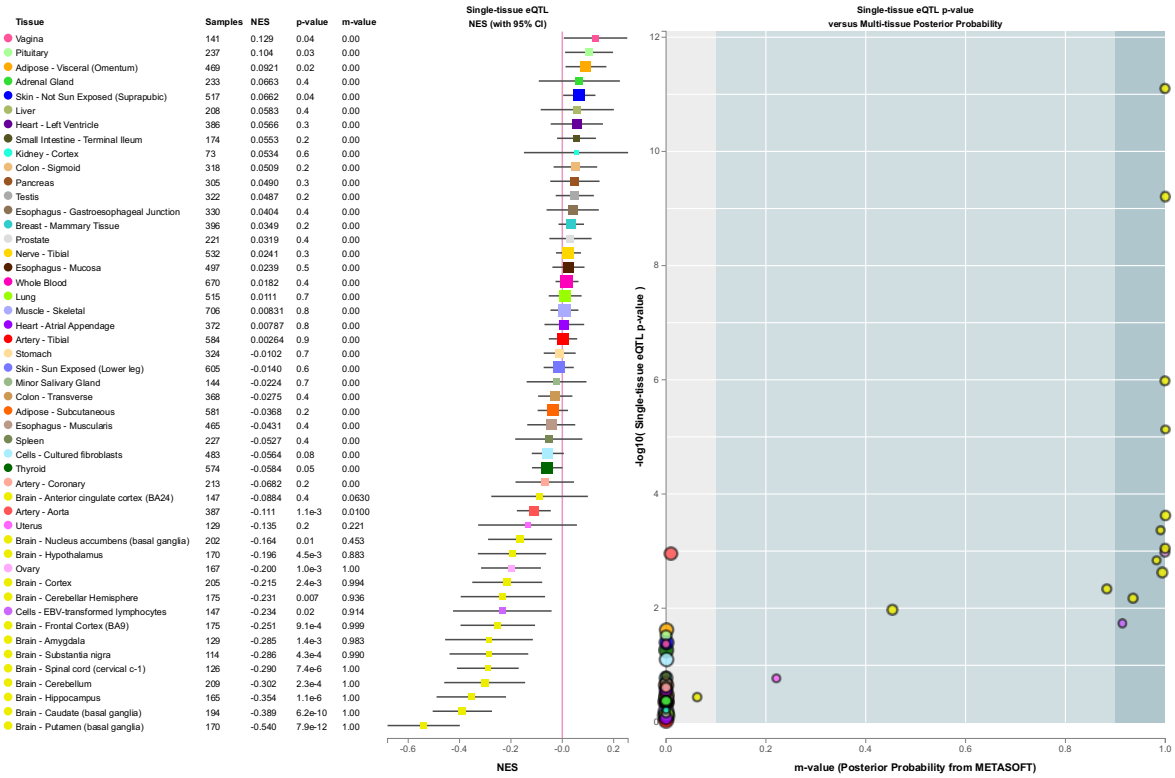
Figure S7. eQTL annotations for rs13132853.



Tissue	Samples	NES	p-value	m-value
Pituitary	237	0.172	0.02	0.175
Uterus	129	-0.00561	0.9	0.786
Adrenal Gland	233	-0.00985	0.9	0.784
Ovary	167	-0.0214	0.8	0.894
Thyroid	574	-0.0257	0.4	0.518
Liver	208	-0.0319	0.5	0.886
Colon - Sigmoid	318	-0.0448	0.3	0.829
Pancreas	305	-0.0496	0.4	0.903
Heart - Atrial Appendage	372	-0.0520	0.2	0.965
Small Intestine - Terminal Ileum	174	-0.0553	0.4	0.909
Brain - Amygdala	129	-0.0673	0.5	0.965
Artery - Coronary	213	-0.0690	0.2	0.977
Esophagus - Mucosa	497	-0.0692	0.05	0.948
Prostate	221	-0.0714	0.2	1.00
Colon - Transverse	368	-0.0758	0.09	0.946
Adipose - Visceral (Omentum)	469	-0.0772	0.05	0.993
Lung	515	-0.0782	0.03	0.983
Cells - Cultured fibroblasts	483	-0.0887	0.01	1.00
Breast - Mammary Tissue	396	-0.0892	0.03	1.00
Nerve - Tibial	532	-0.0913	0.01	1.00
Vagina	141	-0.0933	0.1	0.965
Esophagus - Muscularis	465	-0.0933	0.01	1.00
Brain - Frontal Cortex (BA9)	175	-0.0986	0.2	0.973
Brain - Nucleus accumbens (basal ganglia)	202	-0.106	0.1	0.996
Adipose - Subcutaneous	581	-0.114	0.006	1.00
Artery - Aorta	387	-0.114	3.9e-3	1.00
Spleen	227	-0.114	0.1	0.966
Brain - Anterior cingulate cortex (BA24)	147	-0.123	0.1	0.987
Brain - Hypothalamus	170	-0.124	0.09	0.950
Minor Salivary Gland	144	-0.129	0.2	0.978
Brain - Cortex	205	-0.129	0.09	0.987
Heart - Left Ventricle	386	-0.136	9.9e-4	1.00
Skin - Sun Exposed (Lower leg)	605	-0.136	6.8e-4	1.00
Brain - Hippocampus	165	-0.143	0.1	0.988
Testis	322	-0.144	3.5e-3	1.00
Esophagus - Gastroesophageal Junction	330	-0.147	2.2e-3	1.00
Artery - Tibial	584	-0.153	1.4e-5	1.00
Skin - Not Sun Exposed (Suprapubic)	517	-0.155	5.4e-4	1.00
Muscle - Skeletal	706	-0.158	1.4e-4	1.00
Brain - Caudate (basal ganglia)	194	-0.161	0.01	1.00
Brain - Substantia nigra	114	-0.162	0.1	0.982
Whole Blood	670	-0.168	9.1e-6	1.00
Cells - EBV-transformed lymphocytes	147	-0.168	0.1	0.982
Stomach	324	-0.177	5.9e-4	1.00
Kidney - Cortex	73	-0.191	0.08	0.974
Brain - Spinal cord (cervical c-1)	126	-0.198	0.04	0.993
Brain - Cerebellar Hemisphere	175	-0.236	0.006	1.00
Brain - Putamen (basal ganglia)	170	-0.308	6.4e-5	1.00
Brain - Cerebellum	209	-0.360	3.2e-5	1.00

The SNP rs13132853 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *KLF3*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL. Supplementary Tables XXX include numeric values for other eGenes.

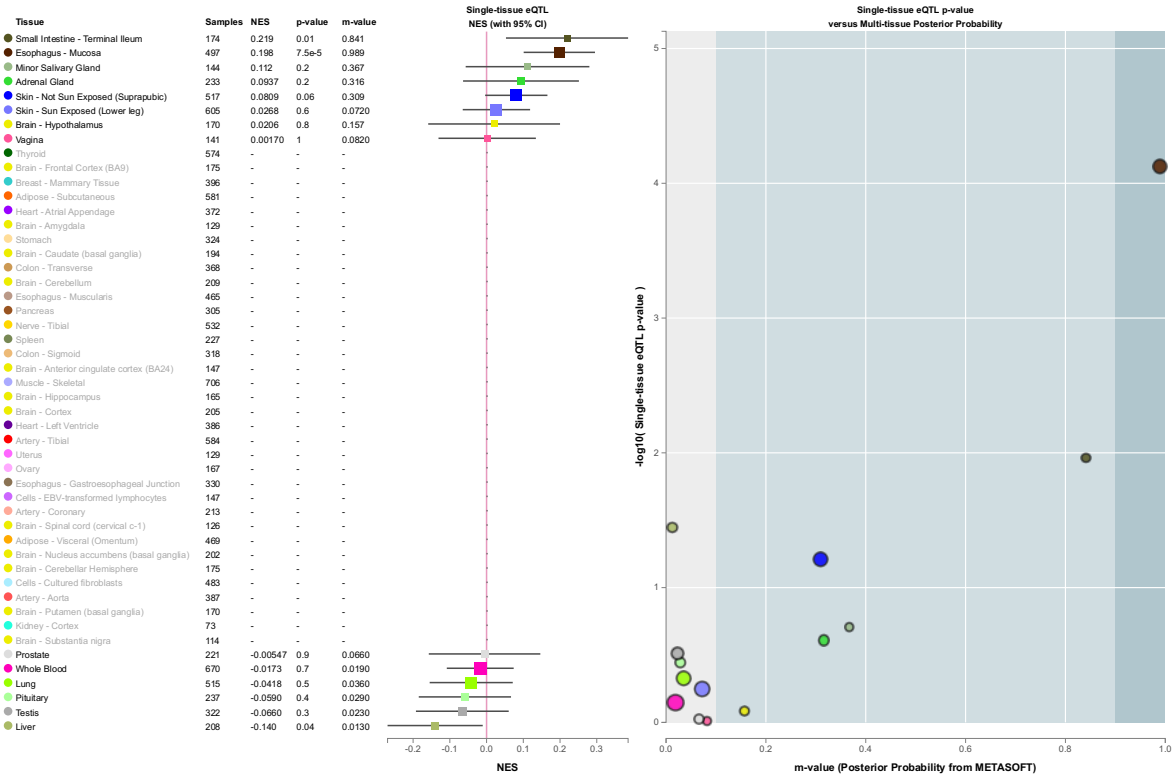
Figure S8. eQTL annotations for rs2790102.

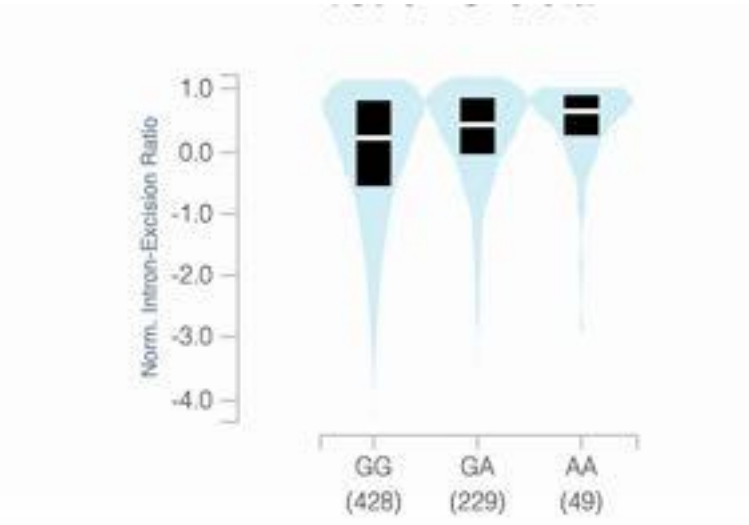


Tissue	Samples	NES	p-value	m-value
Vagina	141	0.129	0.04	0.00
Pituitary	237	0.104	0.03	0.00
Adipose - Visceral (Omentum)	469	0.0921	0.02	0.00
Adrenal Gland	233	0.0663	0.4	0.00
Skin - Not Sun Exposed (Suprapubic)	517	0.0662	0.04	0.00
Liver	208	0.0583	0.4	0.00
Heart - Left Ventricle	386	0.0566	0.3	0.00
Small Intestine - Terminal Ileum	174	0.0553	0.2	0.00
Kidney - Cortex	73	0.0534	0.6	0.00
Colon - Sigmoid	318	0.0509	0.2	0.00
Pancreas	305	0.0490	0.3	0.00
Testis	322	0.0487	0.2	0.00
Esophagus - Gastroesophageal Junction	330	0.0404	0.4	0.00
Breast - Mammary Tissue	396	0.0349	0.2	0.00
Prostate	221	0.0319	0.4	0.00
Nerve - Tibial	532	0.0241	0.3	0.00
Esophagus - Mucosa	497	0.0239	0.5	0.00
Whole Blood	670	0.0182	0.4	0.00
Lung	515	0.0111	0.7	0.00
Muscle - Skeletal	706	0.00831	0.8	0.00
Heart - Atrial Appendage	372	0.00787	0.8	0.00
Artery - Tibial	584	0.00264	0.9	0.00
Stomach	324	-0.0102	0.7	0.00
Skin - Sun Exposed (Lower leg)	605	-0.0140	0.6	0.00
Minor Salivary Gland	144	-0.0224	0.7	0.00
Colon - Transverse	368	-0.0275	0.4	0.00
Adipose - Subcutaneous	581	-0.0368	0.2	0.00
Esophagus - Muscularis	465	-0.0431	0.4	0.00
Spleen	227	-0.0527	0.4	0.00
Cells - Cultured fibroblasts	483	-0.0564	0.08	0.00
Thyroid	574	-0.0584	0.05	0.00
Artery - Coronary	213	-0.0682	0.2	0.00
Brain - Anterior cingulate cortex (BA24)	147	-0.0884	0.4	0.0630
Artery - Aorta	387	-0.111	1.1e-3	0.0100
Uterus	129	-0.135	0.2	0.221
Brain - Nucleus accumbens (basal ganglia)	202	-0.164	0.01	0.453
Brain - Hypothalamus	170	-0.196	4.5e-3	0.883
Ovary	167	-0.200	1.0e-3	1.00
Brain - Cortex	205	-0.215	2.4e-3	0.994
Brain - Cerebellar Hemisphere	175	-0.231	0.007	0.936
Cells - EBV-transformed lymphocytes	147	-0.234	0.02	0.914
Brain - Frontal Cortex (BA9)	175	-0.251	9.1e-4	0.999
Brain - Amygdala	129	-0.285	1.4e-3	0.983
Brain - Substantia nigra	114	-0.286	4.3e-4	0.990
Brain - Spinal cord (cervical c-1)	126	-0.290	7.4e-6	1.00
Brain - Cerebellum	209	-0.302	2.3e-4	1.00
Brain - Hippocampus	165	-0.354	1.1e-6	1.00
Brain - Caudate (basal ganglia)	194	-0.389	6.2e-10	1.00
Brain - Putamen (basal ganglia)	170	-0.540	7.9e-12	1.00

The SNP rs2790102 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *RUNX2*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL. Supplementary Tables XXX include numeric values for other eGenes.

Figure S9. eQTL and sQTL annotations for rs7461069.

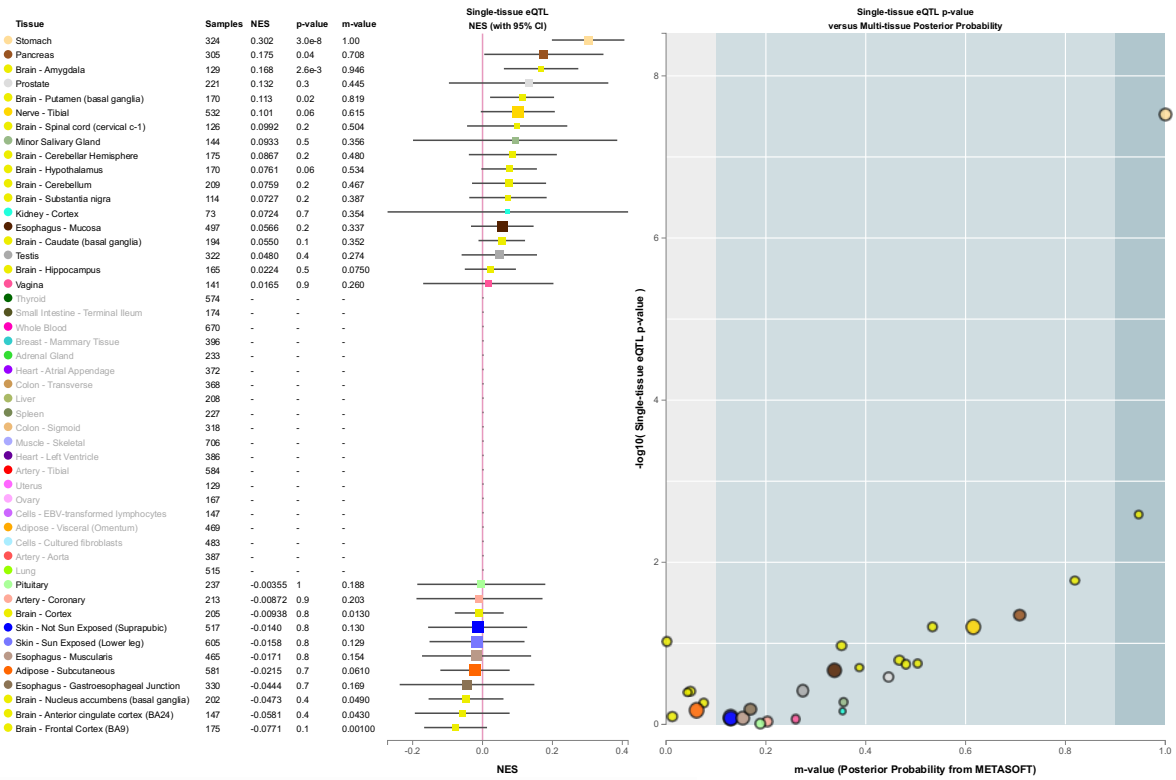


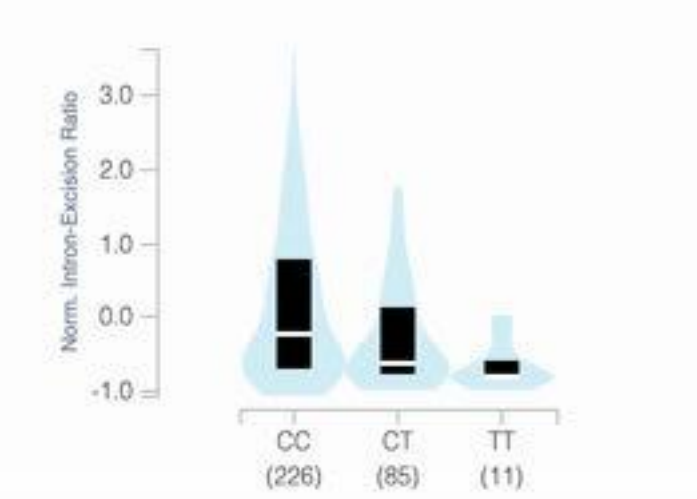


Top: The SNP rs2790102 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *FAM83A-SA1*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL.

Bottom: The SNP rs2790102 was annotated as splicing QTL to the expression of KLHL38 in the Muscle-skeletal tissue (p=9.7x10-6, NES=0.26).

Figure S10. eQTL and sQTL annotations for rs4880424.

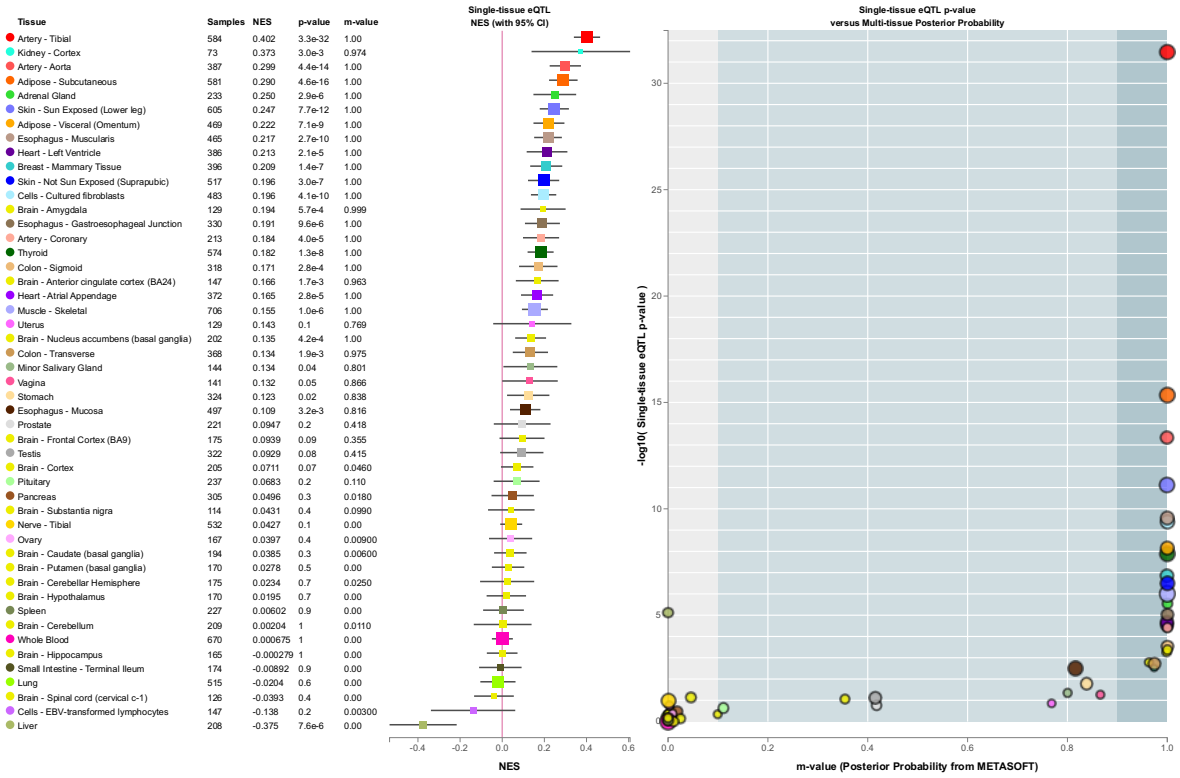




Top: The SNP rs4880424 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *NKX6-2*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL.

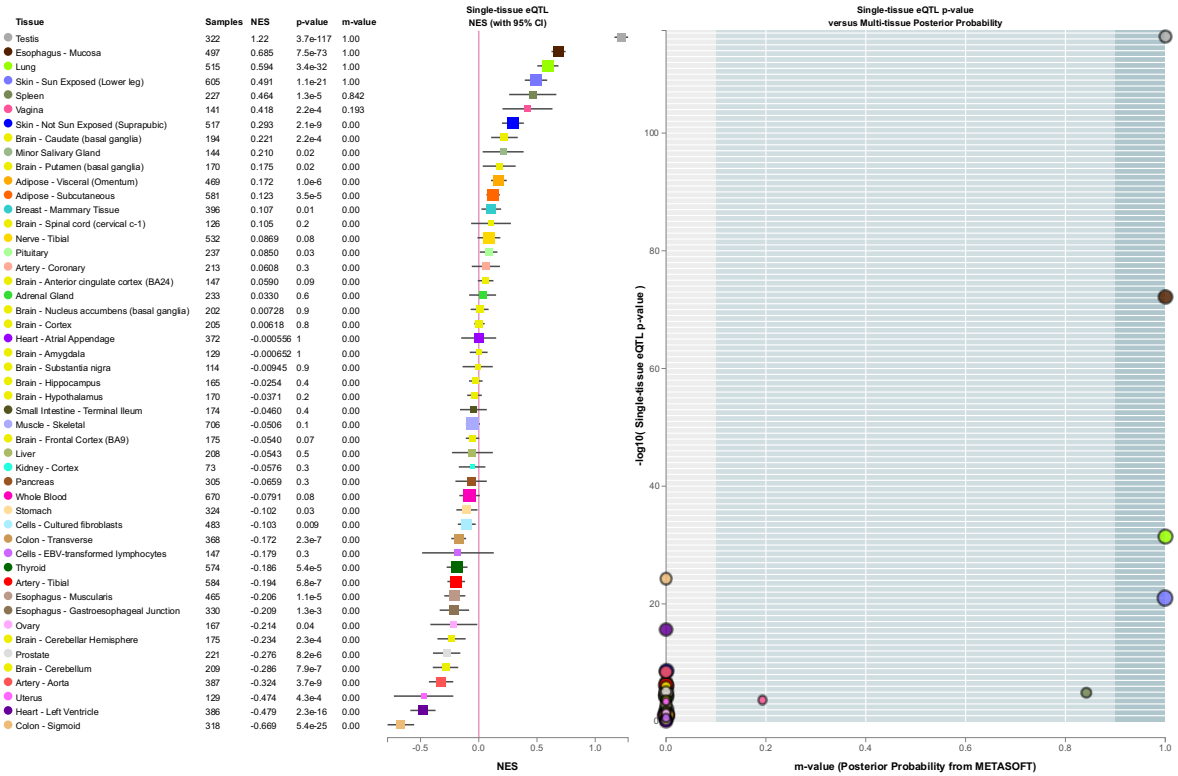
Bottom: The SNP rs4880424 was annotated as splicing QTL to the expression of INPP5A in the Testis tissue (p=6.0x10-7, NES=-0.48).

Figure S11. eQTL and sQTL annotations for rs17203398.



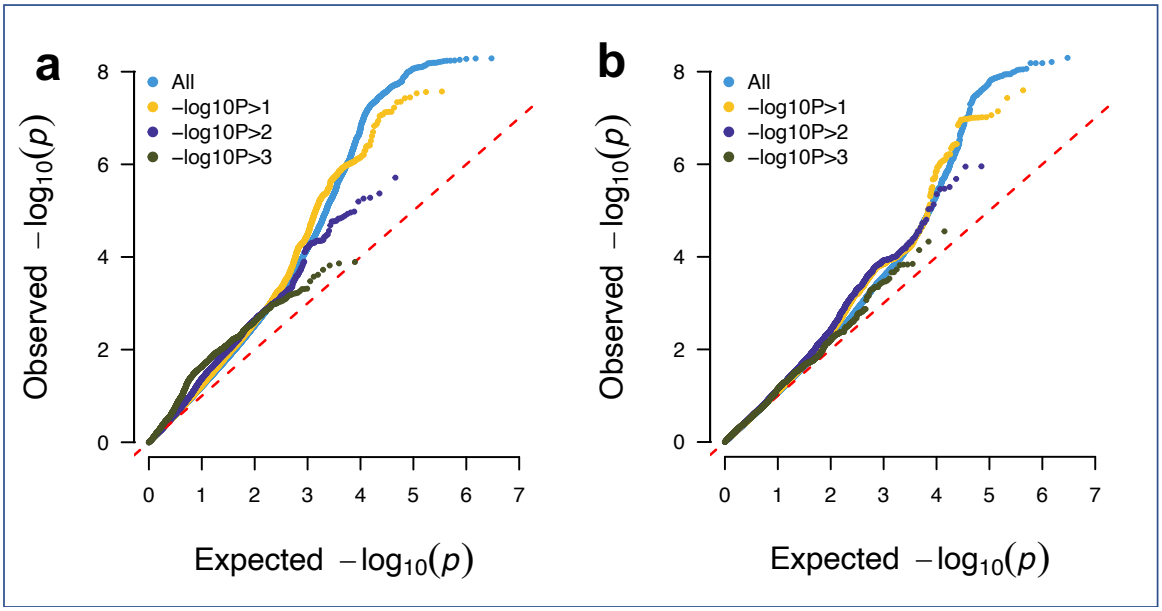
Top: The SNP rs17203398 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *GALC*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL. Supplementary Tables XXX include numeric values for other eGenes.
Bottom: The SNP rs17203398 was annotated as splicing QTL to the gene *GALC* in the Lung tissue (Left, p=2.7x10-8, NES=-0.27), and the non-coding RNA LINC01146 in the liver tissue (Right, p=3.6x10-7, NES=-0.46).

Figure S12. eQTL annotations for rs2106786.



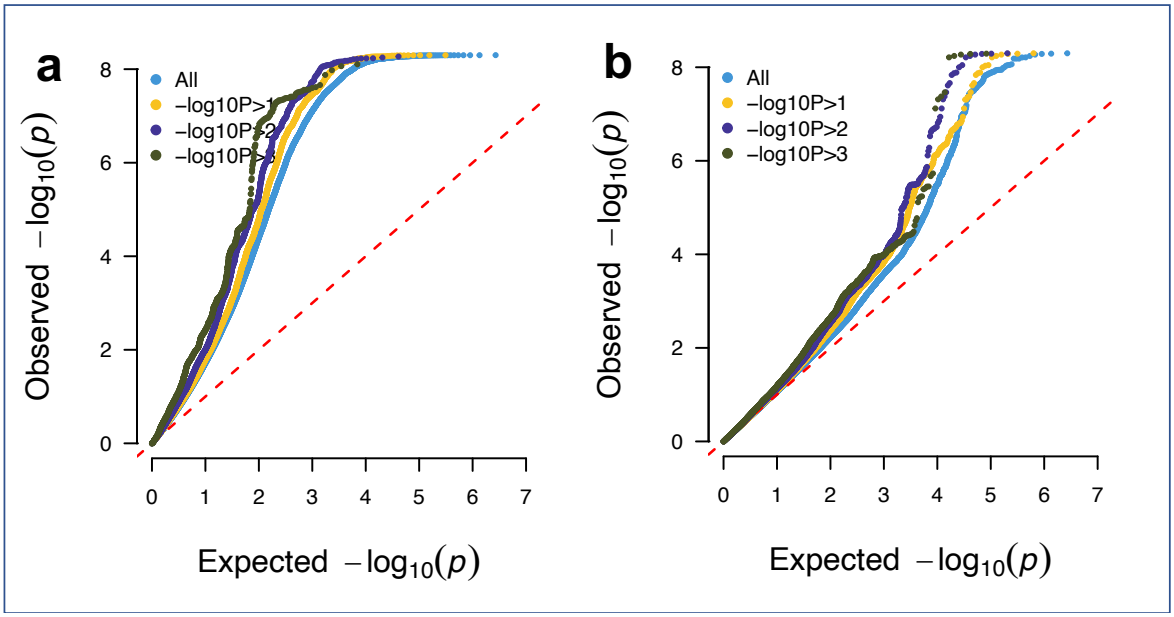
The SNP rs17203398 was search for in the GTEx v8 portal to verify whether it affects the expression level of the nearby gene, *MAPT*, in 49 tissues. The tissue types (Tissue), sample sizes (Samples), normalized effect sizes (NES), eQTL pvalues (p-value), and posterior probability that the SNP affect gene expression estimated by METASOFT[9] (m-value) and 95% CI were shown. Colors indicate tissue types. P-value <1.3x10-4 (0.05 / (49*8) is considered as the existence of statistical evidence; m-value >0.9 is additionally considered as the existence of eQTL. Supplementary Tables XXX include numeric values for other eGenes.

Figure S13. Conditional QQ plot for MDD vs BAG.



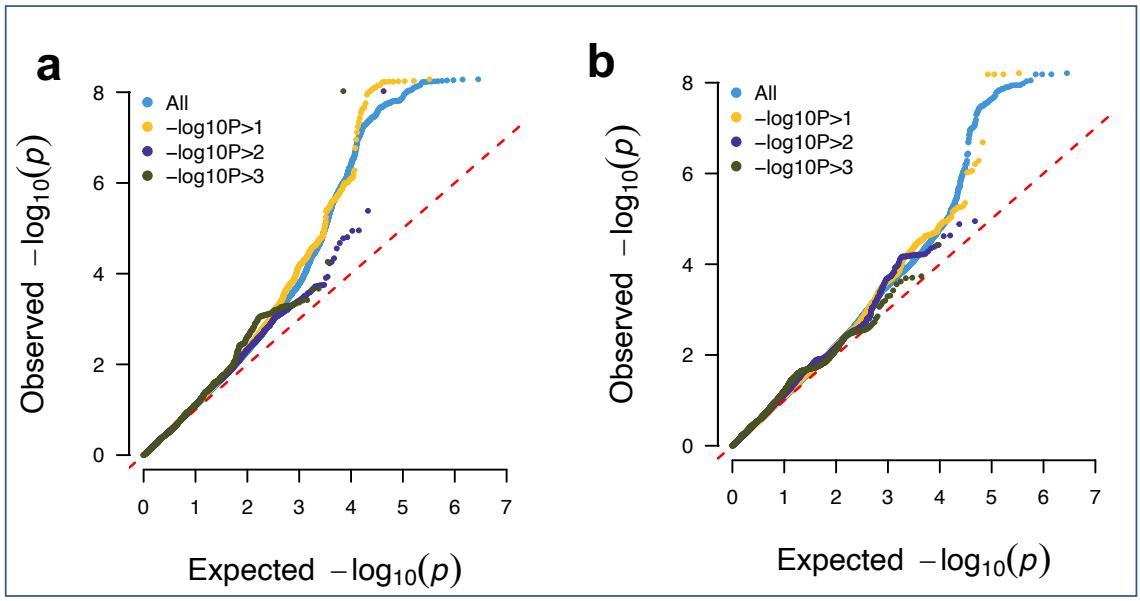
The conditional QQ plots for MDD condition on the levels of association for BAG (**a**), and for BAG condition on the level of association for MDD (**b**). As the association strength to the conditioned disorder increase, a successive leftward deflation in these curves indicates polygenic enrichment.

Figure S14. Conditional QQ plot for SCZ vs BAG.



The conditional QQ plots for SCZ condition on the levels of association for BAG (**a**), and for BAG condition on the level of association for SCZ (**b**). As the association strength to the conditioned disorder increase, a successive leftward deflation in these curves indicates polygenic enrichment.

Figure S15. Conditional QQ plot for PD vs BAG excluding chr17.



The conditional QQ plots for PD condition on the levels of association for BAG (**a**), and for BAG condition on the level of association for PD (**b**). As the association strength to the conditioned disorder increase, a successive leftward deflation in these curves indicates polygenic enrichment.

Table S1. Non-UK Biobank model training datasets.

Name	Size	Subset	Ages	Train
Autism Brain Imaging Data Exchange	555	Healthy Controls	6-56	356
Autism Brain Imaging Data Exchange II	184	Healthy Controls	8-64	118
ADHD 200	606	Healthy Controls	7-26	388
Amsterdam Open MRI Collection	819	ID1000	19-26	524
Beijing Normal University Enhanced Sample	180	Full	17-28	116
Cambridge Center for Aging and Neuroscience	653	Full	18-89	418
Consortium for Reliability and Reproducibility	1368	Full	6-84	880
Dallas Lifespan Brain Study	314	Full	21-89	201
ds000119	73	Full	8-27	47
ds000202	95	Full	18-30	61
ds000222	79	Full	21-73	51
1000 Functional Connectomes Project	812	Full	8-78	520
Healthy Brain Network	1855	Full	5-22	1188
Human Connectome Project	1113	Full	22-37	712
Max Planck Institute Leipzig Mind-Brain-Body	73	Full	22-68	47
Enhanced Nathan Kline Institute - Rockland Sample	928	Full	6-85	594
Open Access Series of Imaging Studies 3	1264	Healthy Controls	42-95	817
Pediatric, Imaging and Neurocognition	1174	Full	3-21	752
Southwest University Adult Lifespan Dataset	494	Full	19-80	316
Southwest University Longitudinal Imaging Multimodal	573	Full	17-27	367
Total	**13212**	**-**	**3-95**	**8473**

The dataset was compiled from 20 sources, spanning a wide age range and a multitude of scanners and scanning protocols. The number of participants from each source used for training is shown in the respective columns.

Table S2. Characteristics for the five folds data from UK Biobank.

Index	N	Age range (years)	Sex (F)	MAE(years)
Fold 0	5622	40-70	2907	2.49
Fold 1	5601	40-70	2885	2.44
Fold 2	5601	40-70	2888	2.45
Fold 3	5650	40-70	2914	2.44
Fold 4	5630	40-70	2907	2.44

Four folds were used for model training and tuning, and the trained model was used for predicting brain age for the hold-off fold. This procedure was iterated five times until all subjects have a brain age estimated from independent training models. Indexes corresponding to those used in Supplementary Figure 1; N, sample size; MAE, mean absolute error for predicted brain age.

Table S3. Statistics for lead SNPs of each associated region in each fold.

Fold	SNP	BP	A1	t-score	P
fold0	rs2790102	45432214	A	-2.54	1.12e-2
fold1	rs2790102	45432214	A	-1.95	5.09e-2
fold2	rs2790102	45432214	A	-3.56	3.77e-4
fold3	rs2790102	45432214	A	-1.86	6.36e-2
fold4	rs2790102	45432214	A	-2.97	2.98e-3
fold0	rs4880424	134584577	T	3.76	1.7e-4
fold1	rs4880424	134584577	T	3.16	1.61e-3
fold2	rs4880424	134584577	T	3.12	1.79e-3
fold3	rs4880424	134584577	T	-0.54	0.59
fold4	rs4880424	134584577	T	2.75	5.97e-3
fold0	rs17203398	88449847	C	-4.13	3.71e-5
fold1	rs17203398	88449847	C	-1.37	0.17
fold2	rs17203398	88449847	C	-4.35	1.41e-5
fold3	rs17203398	88449847	C	-2.04	4.19e-2
fold4	rs17203398	88449847	C	-2.44	1.47e-2
fold0	rs2106786	43919096	G	3.72	2.05e-4
fold1	rs2106786	43919096	G	4.46	8.42e-6
fold2	rs2106786	43919096	G	4.45	8.59e-6
fold3	rs2106786	43919096	G	5.50	4.01e-8
fold4	rs2106786	43919096	G	4.22	2.52e-5
fold0	rs73185796	183975709	T	-1.21	0.23
fold1	rs73185796	183975709	T	-2.84	4.52e-3
fold2	rs73185796	183975709	T	-3.39	6.93e-4
fold3	rs73185796	183975709	T	-2.32	2.02e-2
fold4	rs73185796	183975709	T	-2.72	6.57e-3
fold0	rs13132853	38680015	G	3.59	3.32e-4
fold1	rs13132853	38680015	G	4.18	2.96e-5
fold2	rs13132853	38680015	G	4.29	1.82e-5
fold3	rs13132853	38680015	G	4.10	4.16e-5
fold4	rs13132853	38680015	G	3.378	7.35e-4
fold0	rs79107704	78388694	A	2.26	2.38e-2
fold1	rs79107704	78388694	A	1.64	0.10
fold2	rs79107704	78388694	A	3.27	1.1e-3
fold3	rs79107704	78388694	A	2.92	3.57e-3
fold4	rs79107704	78388694	A	2.49	1.29e-2
fold0	rs7461069	124669029	A	-3.10	1.94e-3
fold1	rs7461069	124669029	A	-2.64	8.37e-3
fold2	rs7461069	124669029	A	-1.47	0.14
fold3	rs7461069	124669029	A	-2.04	4.13e-2
fold4	rs7461069	124669029	A	-3.39	7.12e-4

BP: base pair position on the hg19 build; A1: effective allele; t-score: effect size divided by standard error; P: association p values. Numeric values were rounded to two decimal places.

Table S4. Statistics for lead SNPs of each associated region in each replication samples.

a.

CHR	BP	SNP	A1	N	t-score	P
4	38680015	rs13132853	G	702	0.26	0.79
6	45432214	rs2790102	A	702	-0.13	0.90
8	124669029	rs7461069	A	702	-2.30	2.19e-2
10	134584577	rs4880424	T	702	-0.72	0.47
14	88449847	rs17203398	C	702	-2.70	7.17e-3
17	43919096	rs2106786	G	702	0.86	0.39

b.

CHR	BP	SNP	A1	N	t-score	P
3	183975709	rs73185796	T	321	0.88	0.38
4	38680015	rs13132853	G	321	-0.44	0.66
6	45432214	rs2790102	A	321	-1.06	0.29
8	124669029	rs7461069	A	321	0.19	0.85
10	134584577	rs4880424	T	321	-0.19	0.85
14	88449847	rs17203398	C	321	0.16	0.87
17	43919096	rs2106786	G	321	0.52	0.60

c.

CHR	BP	SNP	A1	N	t-score	P
3	183975709	rs73185796	T	608	-1.73	8.50e-2
4	38680015	rs13132853	G	600	-0.44	0.66
5	78388694	rs79107704	A	612	1.11	0.27
6	45432214	rs2790102	A	501	-1.22	0.22
8	124669029	rs7461069	A	595	1.004	0.32
10	134584577	rs4880424	T	602	0.35	0.72
14	88449847	rs17203398	C	611	0.66	0.51

a. ADNI I sample; **b.** ADNI II sample; **c.** Local healthy control sample.

The ADNI samples I and II were impudently imputed on two the 1000 Genomes Project Phases3 reference data using the software Genipe[3].

CHR: chromosome number; BP: base pair position on the hg19 build; A1: effective allele; N, number of subjects analyzed; t-score: effect size divided by standard error; P: association p values. Numeric values were rounded to two decimal places.

Table S5 Function annotation for genes in Figure 1c.

1. Symbol: **BHMT2**; Name: **Betaine--Homocysteine S-Methyltransferase 2**

Homocysteine is a sulfur-containing amino acid that plays a crucial role in methylation reactions. Transfer of the methyl group from betaine to homocysteine creates methionine, which donates the methyl group to methylate DNA, proteins, lipids, and other intracellular metabolites. The protein encoded by this gene is one of two methyl transferases that can catalyze the transfer of the methyl group from betaine to homocysteine. Anomalies in homocysteine metabolism have been implicated in disorders ranging from vascular disease to neural tube birth defects such as spina bifida, Orofacial Cleft. Among its related pathways are Metabolism and Sulfur amino acid metabolism.

2. Symbol: **AP2M1**; Name: **Adaptor Related Protein Complex 2 Subunit Mu 1**

This gene encodes a subunit of the heterotetrameric coat assembly protein complex 2 (AP2), which belongs to the adaptor complexes medium subunits family. The encoded protein is required for the activity of a vacuolar ATPase, which is responsible for proton pumping occurring in the acidification of endosomes and lysosomes. The encoded protein may also play an important role in regulating the intracellular trafficking and function of CTLA-4 protein. Among its related pathways are Arf1 pathway and Metabolism. Adaptor protein complexes function in protein transport via transport vesicles in different membrane traffic pathways. AP-2 is involved in clathrin-dependent endocytosis in which cargo proteins are incorporated into vesicles surrounded by clathrin (clathrin-coated vesicles, CCVs) which are destined for fusion with the early endosome. Clathrin-associated adaptor protein (AP) complexes which can bind directly to both the clathrin lattice and to the lipid and protein components of membranes are considered to be the major clathrin adaptors contributing the CCV formation. AP-2 may also play a role in maintaining normal post-endocytic trafficking through the ARF6-regulated, non-clathrin pathway. During long-term potentiation in hippocampal neurons, AP-2 is responsible for the endocytosis of ADAM10. Diseases associated with AP2M1 include *Intellectual Developmental Disorder*, *Autosomal Dominant 60*, *Seizures* and *Epilepsy*, *Myoclonic-Atonic Seizures*.

3. Symbol: **PSMD2**; Name: **Proteasome 26S Subunit Ubiquitin Receptor, Non-ATPase 2**

The 26S proteasome is a multicatalytic proteinase complex with a highly ordered structure composed of 2 complexes, a 20S core and a 19S regulator. Proteasomes are distributed throughout eukaryotic cells at a high concentration and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. An essential function of a modified proteasome, the immunoproteasome, is the processing of class I MHC peptides. This gene encodes one of the non-ATPase subunits of the 19S regulator lid. In addition to participation in proteasome function, this subunit may also participate in the TNF signalling pathway since it interacts with the tumor necrosis factor type 1 receptor. This complex plays a key role in the maintenance of protein homeostasis by removing misfolded or damaged proteins, which could impair cellular functions, and by removing proteins whose functions are no longer required. Therefore, the proteasome participates in numerous cellular processes, including cell cycle progression, apoptosis, or DNA damage repair.

4. Symbol: **HEXIM1**; Name: **HEXIM P-TEFb Complex Subunit 1**

Expression of this gene is induced by hexamethylene-bis-acetamide in vascular smooth muscle cells. Among its related pathways are Initiation of transcription and translation elongation at the HIV-1 LTR and Chromatin Regulation/Acetylation. It is a transcriptional regulator which functions as a general RNA polymerase II transcription inhibitor. Core component of the 7SK RNP complex: in cooperation with 7SK snRNA sequesters P-TEFb in a large inactive 7SK snRNP complex preventing RNA polymerase II phosphorylation and subsequent transcriptional elongation. It may also regulate NF-kappa-B, ESR1, NR3C1 and CIITA-dependent transcriptional and plays a role in the regulation of DNA virus-mediated innate immune response by assembling into the HDP-RNP complex, a complex that serves as a platform for IRF3 phosphorylation and subsequent innate immune response activation through the cGAS-STING pathway. Diseases associated with HEXIM1 include *Nut Midline Carcinoma* and *Immune Deficiency Disease*.

5. Symbol: **EIF2B5**; Name: **Eukaryotic Translation Initiation Factor 2B Subunit Epsilon**

This gene encodes one of five subunits of eukaryotic translation initiation factor 2B (EIF2B), a GTP exchange factor for eukaryotic initiation factor 2 and an essential regulator for protein synthesis. Among its related pathways are Translational Control and Peptide chain elongation. It catalyzes the exchange of eukaryotic initiation factor 2-bound GDP for GTP. Mutations in this gene and the genes encoding other EIF2B subunits have been associated with *leukoencephalopathy* with vanishing white matter.

6. Symbol: **NMT1**; Name: **N-Myristoyltransferase 1**

Myristate, a rare 14-carbon saturated fatty acid, is cotranslationally attached by an amide linkage to the N-terminal glycine residue of cellular and viral proteins with diverse functions. N-myristoyltransferase catalyzes the transfer of myristate from CoA to proteins. N-myristoylation appears to be irreversible and is required for full expression of the biologic activities of several N-myristoylated proteins, including the alpha subunit of the signal-transducing guanine nucleotide-binding protein (G protein). Among its related pathways are Activation of BH3-only proteins and HIV Life Cycle. It adds a myristoyl group to the N-terminal glycine residue of certain cellular and viral proteins. Diseases associated with NMT1 include *Gallbladder Cancer* and *Noonan Syndrome 11*.

7. Symbol: **DVL3**; Name: **Dishevelled Segment Polarity Protein 3**

This gene is a member of a multi-gene family which shares strong similarity with the Drosophila dishevelled gene, dsh. The Drosophila dishevelled gene encodes a cytoplasmic phosphoprotein that regulates cell proliferation. It also Involved in the signal transduction pathway mediated by multiple Wnt genes.

8. Symbol: **PARL**; Name: **Presenilin Associated Rhomboid Like**

This gene encodes a member of the rhomboid family of intramembrane serine proteases that is localized to the inner mitochondrial membrane. The encoded protein regulates mitochondrial remodeling and apoptosis through regulated substrate proteolysis. Proteolytic processing of the encoded protein results in the release of a small peptide, P-beta, which may transit to the nucleus. This protein is required for the control of apoptosis during postnatal growth and promotes changes in mitochondria morphology regulated by phosphorylation of P-beta domain. Mutations in this gene may be associated with *Parkinson's disease*.

9. Symbol: **DCAKD**; Name: **Dephospho-CoA Kinase Domain Containing**

The protein coded by this gene is predicted to enable dephospho-CoA kinase activity and involved in coenzyme A biosynthetic process. Gene Ontology (GO) annotations related to this gene include dephospho-CoA kinase activity.

10. Symbol: **INPP5A**; Name: **Inositol Polyphosphate-5-Phosphatase A**

The protein encoded by this gene is a membrane-associated type I inositol 1,4,5-trisphosphate (InsP3) 5-phosphatase. InsP3 5-phosphatases hydrolyze Ins(1,4,5)P3, which mobilizes intracellular calcium and acts as a second messenger mediating cell responses to various stimulation. Among its related pathways are Metabolism and superpathway of D-myo-inositol (1,4,5)-trisphosphate metabolism. Diseases associated with INPP5A include *Alternating Esotropia* and *Lowe Oculocerebrorenal Syndrome*.

11. Symbol: **FAM114A1**; Name: **Family With Sequence Similarity 114 Member A1**

The protein encoded by this gene belongs to the FAM114 family and may play a role in *neuronal cell development*. It is a phosphatase that specifically hydrolyzes the 5-phosphate of inositol 1,4,5-trisphosphate to inositol 1,4-bisphosphate, and inositol 1,3,4,5-tetrasphosphate to inositol 1,3,4-trisphosphate. It also may play a crucial role in the survival of cerebellar Purkinje cells.

12. Symbol: **ALG3**; Name: **ALG3 Alpha-1,3- Mannosyltransferase**

This gene encodes a member of the ALG3 family. The encoded protein catalyses the addition of the first dol-P-Man derived mannose in an alpha 1,3 linkage to Man5GlcNAc2-PP-Dol. Defects in this gene have been associated with c*ongenital disorder of glycosylation type Id* (CDG-Id) characterized by abnormal N-glycosylation, and Congenital Disorders Of N-Linked Glycosylation. Among its related pathways are Synthesis of substrates in N-glycan biosythesis and Metabolism of proteins.

14. Symbol: **KLF3**; Name: **Kruppel Like Factor 3**

The protein coded by this gene enables sequence-specific double-stranded DNA binding activity. It is predicted to be involved in regulation of transcription by RNA polymerase II. The protein binds to the CACCC box of erythroid cell-expressed genes and may play a role in hematopoiesis.

15. Symbol: **ARHGAP27**; Name: **Rho GTPase Activating Protein 27**

This gene encodes a member of a large family of proteins that activate Rho-type guanosine triphosphate (GTP) metabolizing enzymes. The encoded protein may pay a role in clathrin-mediated endocytosis. GTPase activators for the Rho-type GTPases act by converting them to an inactive GDP-bound state.

16. Symbol: **ABCC5**; Name: **ATP Binding Cassette Subfamily C Member 5**

The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes. This protein is a member of the MRP subfamily which is involved in multi-drug resistance. It functions in the cellular export of its substrate, cyclic nucleotides. This export contributes to the degradation of phosphodiesterases and possibly an elimination pathway for cyclic nucleotides. Studies show that this protein provides resistance to thiopurine anticancer drugs, 6-mercatopurine and thioguanine, and the *anti-HIV drug 9*-(2-phosphonylmethoxyethyl) adenine. Among its related pathways are Metabolism and Glycosaminoglycan metabolism. It acts as a heme transporter required for the translocation of cytosolic heme to the secretory pathway and may play a role in energy metabolism by regulating the glucagon-like peptide 1 (GLP-1) secretion from enteroendocrine cells. This protein may be involved in resistance to thiopurines in *acute lymphoblastic leukemia* and *antiretroviral* nucleoside analogs in HIV-infected patients. Diseases associated with ABCC5 include *Primary Angle-Closure Glaucoma* and *Episodic Kinesigenic Dyskinesia 1*.

17. Symbol: **NSF**; Name: **N-Ethylmaleimide Sensitive Factor, Vesicle Fusing ATPase**

The protein coded by this gene enables PDZ domain binding activity and ionotropic glutamate receptor binding activity. It is involved in intracellular protein transport, positive regulation of protein catabolic process, and positive regulation of receptor recycling. Among its related pathways are Neuroscience and Vesicle-mediated transport. Diseases associated with NSF include *Developmental And Epileptic Encephalopathy 96* and *Tetanus*.

18. Symbol: **FMNL1**; Name: **Formin Like 1**

This gene encodes a formin-related protein. Formin-related proteins have been implicated in morphogenesis, cytokinesis, and cell polarity. The protein may play a role in the control of cell motility, survival of macrophages, the regulation of cell morphology and cytoskeletal organization. It is required in the cortical actin filament dynamics and cell shape.

19. Symbol: **ACBD4**; Name: **Acyl-CoA Binding Domain Containing 4**

This gene encodes a member of the acyl-coenzyme A binding domain containing protein family. All family members contain the conserved acyl-Coenzyme A binding domain, which binds acyl-CoA thiol esters. They are thought to play roles in acyl-CoA dependent lipid metabolism by binding to medium- and long-chain acyl-CoA esters and may function as an intracellular carrier of acyl-CoA esters.

20. Symbol: **GALC**; Name: **Galactosylceramidase**

This gene encodes a lysosomal protein which hydrolyzes the galactose ester bonds of galactosylceramide, galactosylsphingosine, lactosylceramide, and monogalactosyldiglyceride. Enzyme with very low activity responsible for the lysosomal catabolism of galactosylceramide, a major lipid in myelin, kidney and epithelial cells of small intestine and colon. Mutations in this gene have been associated with *Krabbe disease*, also known as globoid cell leukodystrophy.

21. Symbol: **KANSL1**; Name: **KAT8 Regulatory NSL Complex Subunit 1**

This gene encodes a nuclear protein that is a subunit of two protein complexes involved with histone acetylation, the MLL1 complex and the NSL1 complex. As part of the NSL complex it is involved in acetylation of nucleosomal histone H4 on several lysine residues and therefore may be involved in the regulation of transcription.

22. Symbol: **PLEKHM1**; Name: **Pleckstrin Homology And RUN Domain Containing M1**
The protein encoded by this gene is essential for bone resorption, and may play a critical role in vesicular transport in the osteoclast. It acts as a multivalent adapter protein that regulates Rab7-dependent and HOPS complex-dependent fusion events in the endolysosomal system and couples autophagic and the endocytic trafficking pathways. PLEKGN1 is a dual effector of RAB7A and ARL8B that simultaneously binds these GTPases, bringing about clustering and fusion of late endosomes and lysosomes. It is required for late stages of endolysosomal maturation, facilitating both endocytosis-mediated degradation of growth factor receptors and autophagosome clearance. For example, in case of infection, it contributes to Salmonella typhimurium pathogenesis by supporting the integrity of the Salmonella-containing vacuole (SCV) probably in concert with the HOPS complex and Rab7. Mutations in this gene are associated with *autosomal recessive osteopetrosis type 6* (OPTB6).
23. Symbol: **FBXO32**; Name: **F-Box Protein 32**
This gene encodes a member of the F-box protein family which is characterized by an approximately 40 amino acid motif, the F-box. The F-box proteins constitute one of the four subunits of the ubiquitin protein ligase complex called SCFs (SKP1-cullin-F-box), which function in phosphorylation-dependent ubiquitination. The protein probably recognizes and binds to phosphorylated target proteins during skeletal muscle atrophy. It is highly expressed during *muscle atrophy*, whereas mice deficient in this gene were found to be resistant to atrophy. This protein is thus a potential drug target for the treatment of *muscle atrophy.*
24. Symbol: **PLCD3**; Name: **1-Phosphatidylinositol 4,5-Bisphosphate Phosphodiesterase Delta-3**
This gene encodes a member of the phospholipase C family, which catalyze the hydrolysis of phosphatidylinositol 4,5-bisphosphate to generate the second messengers diacylglycerol and inositol 1,4,5-trisphosphate (IP3). Diacylglycerol and IP3 mediate a variety of cellular responses to extracellular stimuli by inducing protein kinase C and increasing cytosolic Ca(2+) concentrations. Its activity is inhibited by spermine, sphingosine, and several phospholipids. It is essential for trophoblast and placental development and may participate in cytokinesis by hydrolyzing PIP2 at the cleavage furrow; It also regulates *neurite outgrowth* through the inhibition of RhoA/Rho kinase signaling.

GROUP 2
25. Symbol: **TLR10**; Name: **Toll Like Receptor 10**
The protein encoded by this gene is a member of the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity. TLRs are highly conserved from Drosophila to humans and share structural and functional similarities. They recognize pathogen-associated molecular patterns (PAMPs) that are expressed on infectious agents, and mediate the production of cytokines necessary for the development of *effective immunity*. It acts via MYD88 and TRAF6, leading to NF-kappa-B activation, cytokine secretion and the inflammatory response.
26. Symbol: **GPR65**; Name: **G Protein-Coupled Receptor 65**
The protein encoded by this gene enables G protein-coupled receptor activity and is involved in several processes, including actin cytoskeleton reorganization, activation of GTPase activity, and positive regulation of stress fiber assembly. It is a receptor for the glycosphingolipid psychosine (PSY) and several related glycosphingolipids. It also plays a role in *immune response* by maintaining lysosome function and supporting phagocytosis-mediated intracellular *bacteria clearance* and may have a role in activation-induced cell death or differentiation of T-cells.
27. Symbol: **TLR1**; Name: **Toll Like Receptor 1**
The protein encoded by this gene is a member of the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity. This gene is ubiquitously expressed, and at higher levels than other TLR genes. This protein participates in the innate immune response to microbial agents, specifically recognizes diacylated and triacylated lipopeptides. It cooperates with TLR2 to mediate the innate immune response to bacterial

lipoproteins or lipopeptides by forming the activation cluster TLR2:TLR1:CD14 in response to triacylated lipopeptides. This cluster triggers signaling from the cell surface and subsequently is targeted to the Golgi in a lipid-raft dependent pathway, acting via MYD88 and TRAF6, leading to NF-kappa-B activation, cytokine secretion and the inflammatory response.

28. Symbol: **KLHL38**; Name: **Kelch Like Family Member 38**

KLHL38 is a Protein Coding gene. Diseases associated with KLHL38 include *Posterior Myocardial Infarction.*

29. Symbol: **SPATA32**; Name: **Spermatogenesis Associated 32**

The protein encoded by this gene is predicted to enable actin binding activity, to be involved in spermatogenesis and active in perinuclear region of cytoplasm.

30. Symbol: **C17orf104**

This is an RNA Gene and is affiliated with the lncRNA class; Its function is unclear.

31. Symbol: **SPPL2C**; Name: **Signal Peptide Peptidase Like 2C**

The protein encoded by this gene enables protein homodimerization activity. And, it is predicted to be involved in membrane protein proteolysis. SPPL2C is integral component of cytoplasmic side of endoplasmic reticulum membrane and integral component of lumenal side of endoplasmic reticulum membrane. Diseases associated with SPPL2C include *chromosome 17Q21.31 Duplication Syndrome* and *Caplan's Syndrome*.

32. Symbol: **LRRC37A**; Name: **Leucine Rich Repeat Containing 37A**

The protein encoded by this gene is predicted to be integral component of membrane. Diseases associated with LRRC37A include *Koolen-De Vries Syndrome* and *Supranuclear Palsy, Progressive, 1*.

33. Symbol: **ARL17A**; Name: **ADP Ribosylation Factor Like GTPase 17A**

The protein encoded by this gene is predicted to enable GTP binding activity, involved in intracellular protein transport and vesicle-mediated transport. It functions as an allosteric activator of the cholera toxin catalytic subunit, an ADP-ribosyltransferase. Diseases associated with ARL17A include *Bardet-Biedl Syndrome* and *Retinitis Pigmentosa*.

34. Symbol: **ARL17B**; Name: **ADP Ribosylation Factor Like GTPase 17B**

The protein encoded by this gene is predicted to enable GTP binding activity, involved in intracellular protein transport and vesicle-mediated transport. It functions as an allosteric activator of the cholera toxin catalytic subunit, an ADP-ribosyltransferase.

35. Symbol: **STH**; Name: **Microtubule-Associated Protein Tau (MAPT) Intronic Transcript**

The protein encoded by this gene is involved in positive regulation of mRNA splicing via spliceosome. It is located in nucleus and perinuclear region of cytoplasm. Diseases associated with STH include *Frontotemporal Dementia* and *Chromosome 17Q21.31 Duplication Syndrome*.

36. Symbol: **C10orf91** (**LINC02870**); Name: **Long Intergenic Non-Protein Coding RNA 2870**

No functions have been annotated to this non-coding RNA yet.

37. Symbol: **WNT9B**; Name: **Wnt Family Member 9B**

This gene is a member of the WNT gene family. The WNT gene family consists of structurally related genes that encode secreted signaling proteins. These proteins have been implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis. Study of its expression in the teratocarcinoma cell line NT2 suggests that it may be implicated in the early process of *neuronal* differentiation of NT2 cells induced by retinoic acid. This gene is clustered with WNT3, another family member, in the chromosome 17q21 region. It belongs to the canonical Wnt/beta-catenin signaling pathway. It activates a signaling cascade in the metanephric mesenchyme that induces tubulogenesis and plays a role in cranofacial development and is required for normal fusion of the palate during embryonic.

38. Symbol: **RUNX2**; Name: **RUNX Family Transcription Factor 2**

This gene is a member of the RUNX family of transcription factors and encodes a nuclear protein with an Runt DNA-binding domain. This protein is essential for osteoblastic differentiation and

skeletal morphogenesis and acts as a scaffold for nucleic acids and regulatory factors involved in skeletal gene expression. The protein can bind DNA both as a monomer or, with more affinity, as a subunit of a heterodimeric complex. Two regions of potential trinucleotide repeat expansions are present in the N-terminal region of the encoded protein, and these and other mutations in this gene have been associated with the bone development disorder cleidocranial dysplasia (CCD). Transcript variants that encode different protein isoforms result from the use of alternate promoters as well as alternate splicing. It is essential for the maturation of osteoblasts and both intramembranous and endochondral ossification. In osteoblasts, it supports transcription activation: synergizes with SPEN/MINT to enhance FGFR2-mediated activation of the osteocalcin FGF-responsive element (OCFRE).

39. Symbol: **C17orf53** (HROB); Name: **Homologous Recombination Factor With OB-Fold**

The protein encoded by this gene is predicted to enable single-stranded DNA binding activity and involved in DNA synthesis, DNA repair and interstrand cross-link repair. It functions by recruiting the MCM8-MCM9 helicase complex to sites of DNA damage to promote DNA repair synthesis.

40. Symbol: **EFCAB13**; Name: **EF-Hand Calcium Binding Domain 13**

EFCAB13 is a Protein Coding gene, and its annotations include calcium ion binding.

41. Symbol: **GJC1**; Name: **Gap Junction Protein Gamma 1**

This gene is a member of the connexin gene family. The encoded protein is a component of gap junctions, which are composed of arrays of intercellular channels that provide a route for the diffusion of low molecular weight materials from cell to cell. One gap junction consists of a cluster of closely packed pairs of transmembrane channels, the connexons, through which materials of low MW diffuse from one cell to a neighboring cell. Gene Ontology annotations related to this gene include ion channel activity and gap junction channel activity. They allow passive diffusion of molecules up to 1 kDa, including nutrients, metabolites (glucose), ions (K+, Ca2+) and second messengers (IP3, cAMP).

42. Symbol: **DBF4B**; Name: **DBF4 Zinc Finger B**

This gene encodes a regulator of the cell division cycle 7 homolog (S. cerevisiae) protein, a serine-threonine kinase which links cell cycle regulation to genome duplication. It is a regulatory subunit for CDC7 which activates its kinase activity thereby playing a central role in DNA replication and cell proliferation and is required for progression of S and M phases. The complex CDC7-DBF4B selectively phosphorylates MCM2 subunit at 'Ser-40' and then is involved in regulating the initiation of DNA replication during cell cycle.

43. Symbol: **HEXIM2**; Name: **HEXIM P-TEFb Complex Subunit 2**

This gene encodes a member of the HEXIM family of proteins. This protein is a component of the 7SK small nuclear ribonucleoprotein. This protein has been found to negatively regulate the kinase activity of the cyclin-dependent kinase P-TEFb, which phosphorylates multiple target proteins to promote transcriptional elongation. This gene is located approximately 7 kb downstream from related family member HEXIM1 on chromosome 17. It is a transcriptional regulator which functions as a general RNA polymerase II transcription inhibitor and is a core component of the 7SK RNP complex: in cooperation with 7SK snRNA sequesters P-TEFb in a large inactive 7SK snRNP complex preventing RNA polymerase II phosphorylation and subsequent transcriptional.

44. Symbol: **LRRC37A2**; Name: **leucine rich repeat containing 37 member A2**

The protein encoded by this gene is predicted to be an integral component of membrane. Diseases associated with LRRC37A2 include *Epilepsy*, *Progressive Myoclonic 6* and *Developmental And Epileptic Encephalopathy 96*.

45. Symbol: **DMGDH**; Name: **dimethylglycine dehydrogenase**

This gene encodes an enzyme involved in the catabolism of choline, catalyzing the oxidative demethylation of dimethylglycine to form sarcosine. The enzyme is found as a monomer in the mitochondrial matrix, and uses flavin adenine dinucleotide and folate as cofactors. Mutation in this gene causes *dimethylglycine dehydrogenase deficiency*, characterized by a fishlike body odor, chronic muscle fatigue, and elevated

levels of the muscle form of creatine kinase in serum.

GROUP 3

46. Symbol: **WNT3**; Name: **Wnt family member 3**

The WNT gene family consists of structurally related genes which encode secreted signaling proteins. These proteins have been implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis. Studies of the gene expression suggest that this gene may play a key role in some cases of human breast, rectal, lung, and gastric cancer through activation of the WNT-beta-catenin-TCF signaling pathway. The WNT3 protein functions in the canonical Wnt signaling pathway that results in activation of transcription factors of the TCF/LEF family. And it is required for normal gastrulation, formation of the primitive streak, and for the formation of the mesoderm during early embryogenesis.

47. Symbol: **C1QL1**; Name: **complement C1q like 1**

The protein encoded by this gene is predicted to enable signaling receptor binding activity, act upstream of or within maintenance of synapse structure, motor learning, and *neuron remodeling*. It is predicted to be located in several cellular components, including climbing fiber; presynapse; and synaptic cleft. It may also regulate the number of excitatory synapses that are formed on hippocampus neurons but has no effect on inhibitory synapses.

48. Symbol: **MAPT**; Name: **microtubule associated protein tau**

This gene encodes the microtubule-associated protein tau (MAPT) whose transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species. MAPT transcripts are differentially expressed in the nervous system, depending on stage of neuronal maturation and neuron type. This protein promotes microtubule assembly and stability, and might be involved in the establishment and maintenance of neuronal polarity. The C-terminus binds axonal microtubules while the N-terminus binds neural plasma membrane components, suggesting that tau functions as a linker protein between both. Axonal polarity is predetermined by TAU/MAPT localization (in the neuronal cell) in the domain of the cell body defined by the centrosome. The short isoforms allow plasticity of the cytoskeleton whereas the longer isoforms may preferentially play a role in its stabilization. MAPT gene mutations have been associated with several neurodegenerative disorders such as *Alzheimer's disease*, *Pick's disease*, *frontotemporal dementia*, *cortico-basal degeneration* and *progressive supranuclear palsy*.

49. Symbol: **CRHR1**; Name: **corticotropin releasing hormone receptor 1**

This gene encodes a G-protein coupled receptor that binds neuropeptides of the corticotropin releasing hormone family that are major regulators of the *hypothalamic-pituitary-adrenal pathway*. The encoded protein is essential for the activation of signal transduction pathways that regulate diverse physiological processes including stress, reproduction, immune response and obesity. This G-protein coupled receptor for CRH (corticotropin-releasing factor) and UCN (urocortin) has high affinity for CRH and UCN: Ligand binding causes a conformation change that triggers signaling via guanine nucleotide-binding proteins (G proteins) and down-stream effectors, such as adenylate cyclase. It promotes the activation of adenylate cyclase, leading to increased intracellular cAMP levels, inhibits the activity of the calcium channel CACNA1H. CRHR1 is required for normal embryonic development of the adrenal gland and for normal hormonal responses to stress. 50. Symbol: **ECE2**; Name: **Endothelin Converting Enzyme 2**

The enzym coded by this gene enables metalloendopeptidase activity. It is involved in peptide hormone processing, converts big endothelin-1 to endothelin-1. And it is also involved in the processing of various neuroendocrine peptides, including neurotensin, angiotensin I, substance P, proenkephalin-derived peptides, and prodynorphin-derived peptides.

51. Symbol: **CAMK2N2**; Name: **Calcium/Calmodulin Dependent Protein Kinase II Inhibitor 2**

This gene encodes a protein that is highly similar to the rat CaM-KII inhibitory protein, an inhibitor of calcium/calmodulin-dependent protein kinase II (CAMKII). CAMKII regulates numerous physiological

functions, including neuronal synaptic plasticity through the phosphorylation of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-type glutamate (AMPA) receptors. Studies of the similar protein in rat suggest that this protein may function as a negative regulator of CaM-KII and may act to inhibit the phosphorylation of AMPA receptors. Gene Ontology (GO) annotations related to this gene include protein kinase binding and calcium-dependent protein kinase inhibitor activity. Diseases associated with CAMK2N2 include *Amyotrophic Lateral Sclerosis 21* and *Amyotrophic Lateral Sclerosis Type 6*.

52. Symbol: **VWA5B2**; Name: **Von Willebrand Factor A Domain Containing 5B2**

Variations in this gene have been associated with Mathemtaics ability. Somatic mutations in it have been reported in various cancers.

53. Symbol: **GFAP**; Name: **Glial fibrillary acidic protein**

This gene encodes one of the major intermediate filament proteins of mature astrocytes. It is used as a marker to distinguish astrocytes from other glial cells during development. Mutations in this gene cause Alexander disease, a rare disorder of astrocytes in the central nervous system. GFAP, a class-III intermediate filament, is a cell-specific marker that distinguishes astrocytes from other glial cells during the development of the central nervous system. Defects in GFAP are a cause of *Alexander disease* (ALEXD), which is a rare disorder of the central nervous system.

54. Symbol: **NKX6-2**; Name: **NK6 homeobox 2**

The protein encoded by this gene enables sequence-specific double-stranded DNA binding activity and is Predicted to be involved in cell differentiation, regulation of myelination and transcription. It acts upstream of or within several processes, including negative regulation of transcription by RNA polymerase II, neurogenesis; and neuromuscular process controlling balance. Diseases associated with mutations in this gene include *Spastic Ataxia 8*.

Table S6. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to SCZ MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10137195	14	88444882	C	A	0.36	-0.16	0.0257	28104	1.42e-10	0.44	0.0056	0.0082	73173	0.49
rs2239923	17	43176804	T	C	0.29	-0.15	0.0269	28104	2.18e-8	0.28	0.0251	0.0086	73173	3.33e-3
rs2242189	4	38691024	C	T	0.38	0.21	0.0257	28104	1.24e-16	0.33	-0.0032	0.0084	73173	0.70
rs2790102	6	45432214	A	G	0.35	-0.15	0.0260	28104	8.92e-9	0.32	-0.0063	0.0084	73173	0.45
rs4880424	10	134584577	T	C	0.22	0.16	0.0296	28104	3.69e-8	0.20	-0.0121	0.0098	73173	0.22
rs62063276	17	44036408	G	T	0.21	0.28	0.0293	28104	2.72e-22	0.20	-0.0588	0.0108	60448	5.92e-8
rs7209501	17	43503294	C	A	0.18	0.29	0.0315	28104	5.33e-20	0.16	-0.0482	0.0114	64002	2.50e-5
rs7461069	8	124669029	A	G	0.21	-0.17	0.0309	28104	1.57e-8	0.18	-0.0149	0.01	73173	0.14

Instrumental SNPs select for the BAG to SCZ classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S7. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to BIP MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10137195	14	88444882	C	A	0.36	-0.16	0.0257	28104	1.42e-10	0.39	-0.0151	0.0098	50981	0.13
rs2239923	17	43176804	T	C	0.29	-0.15	0.0269	28104	2.18e-8	0.28	-0.0333	0.0104	50981	1.3e-3
rs2242189	4	38691024	C	T	0.38	0.21	0.0257	28104	1.24e-16	0.34	-0.0021	0.0098	50981	0.83
rs2790102	6	45432214	A	G	0.35	-0.15	0.0260	28104	8.92e-9	0.35	0.0017	0.0098	50981	0.87
rs4880424	10	134584577	T	C	0.22	0.16	0.0296	28104	3.69e-8	0.21	-0.0045	0.0115	50981	0.69
rs62063276	17	44036408	G	T	0.21	0.28	0.0293	28104	2.72e-22	0.19	0.0215	0.0116	50981	0.06
rs7209501	17	43503294	C	A	0.18	0.29	0.0315	28104	5.34e-20	0.16	0.0218	0.0125	50981	0.08
rs7461069	8	124669029	A	G	0.21	-0.17	0.0309	28104	1.57e-8	0.20	0.0134	0.0118	50981	0.26

Instrumental SNPs select for the BAG to BIP classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S8. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to MDD MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10137195	14	88444882	C	A	0.36	-0.16	0.0257	28104	1.421e-10	0.3596	0.0051	0.0082	69115	0.53
rs11079732	17	44161581	T	C	0.45	-0.14	0.0247	28104	4.67e-9	0.4537	-0.0076	0.0092	52547	0.41
rs1879581	17	43545893	C	T	0.18	0.29	0.0315	28104	3.23e-20	0.1766	0.0214	0.0105	69115	0.04
rs2239923	17	43176804	T	C	0.29	-0.15	0.0269	28104	2.18e-08	0.284	0.002	0.0088	69115	0.82
rs2242189	4	38691024	C	T	0.38	0.21	0.0257	28104	1.24e-16	0.3681	-0.0115	0.0082	69115	0.16
rs2790102	6	45432214	A	G	0.35	-0.15	0.0260	28104	8.92e-9	0.3487	0.0041	0.0084	69115	0.62
rs4880424	10	134584577	T	C	0.22	0.16	0.0296	28104	3.69e-8	0.2144	-0.0175	0.0103	69115	0.09
rs7461069	8	124669029	A	G	0.21	-0.17	0.0309	28104	1.57e-8	0.1953	0.0008	0.0099	69115	0.93

Instrumental SNPs select for the BAG to MDD classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S9. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to AD MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10137195	14	88444882	C	A	0.36	-0.1648	0.0257	28104	1.421e-10	0.37	0.0008	0.0022	429826.3	0.71
rs2239923	17	43176804	T	C	0.29	-0.1507	0.0269	28104	4.67e-9	0.28	-0.0032	0.0024	428474.56	0.19
rs2242189	4	38691024	C	T	0.38	0.2126	0.0257	28104	3.23e-20	0.34	-0.0025	0.0023	417305.82	0.27
rs2790102	6	45432214	A	G	0.35	-0.1493	0.0260	28104	2.18e-08	0.35	-0.0056	0.0023	428230.78	0.01
rs4880424	10	134584577	T	C	0.22	0.1632	0.0296	28104	1.24e-16	0.22	0.0003	0.0026	426375.66	0.92
rs62063276	17	44036408	G	T	0.21	0.2842	0.0293	28104	8.92e-9	0.21	-0.0054	0.0027	427814.95	0.04
rs7209501	17	43503294	C	A	0.18	0.2888	0.0315	28104	3.69e-8	0.17	-0.0029	0.0028	429246.56	0.30
rs7461069	8	124669029	A	G	0.21	-0.1745	0.0309	28104	1.57e-8	0.18	-0.0011	0.0028	429105.83	0.69

Instrumental SNPs select for the BAG to AD classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S10. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BAG to PD MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10137195	14	88444882	C	A	0.36	-0.1648	0.0257	28104	1.421e-10	0.37	0.0602	0.0174	62649	5e-4
rs2239923	17	43176804	T	C	0.29	-0.1507	0.0269	28104	4.67e-9	0.31	0.0446	0.0214	62649	0.04
rs2242189	4	38691024	C	T	0.38	0.2126	0.0257	28104	3.23e-20	0.36	0.0151	0.0233	49863	0.52
rs2790102	6	45432214	A	G	0.35	-0.1493	0.0260	28104	2.18e-08	0.34	-0.0111	0.0238	49863	0.64
rs4880424	10	134584577	T	C	0.22	0.1632	0.0296	28104	1.24e-16	0.22	0.0078	0.0274	49863	0.77
rs62063276	17	44036408	G	T	0.21	0.2842	0.0293	28104	8.92e-9	0.22	-0.2507	0.027	60613	1.46e-20
rs7209501	17	43503294	C	A	0.18	0.2888	0.0315	28104	3.69e-8	0.19	-0.2225	0.0274	60613	4.66e-16
rs7461069	8	124669029	A	G	0.21	-0.1745	0.0309	28104	1.57e-8	0.19	0.0252	0.0256	62649	0.33

Instrumental SNPs select for the BAG to PD classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S11. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the SCZ to BAG MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10098869	8	4208761	T	C	0.85	-0.07	0.0108	73173	2.85e-10	0.86	-0.0265	0.0358	28104	0.4
7rs10127983	1	153923276	C	T	0.70	-0.05	0.0084	73173	3.11e-8	0.70	-0.0233	0.0271	28104	0.39
rs10167385	2	37238824	C	T	0.68	-0.05	0.0086	73173	7.94e-9	0.75	0.0046	0.0287	28104	0.87
rs10238960	7	70773271	C	T	0.32	-0.05	0.0084	73173	7.65e-9	0.30	-0.021	0.0270	28104	0.44
rs10243922	7	138599980	G	A	0.73	-0.05	0.0089	73173	1.02e-8	0.69	0.0496	0.0267	28104	0.06
rs1034326	6	30084446	A	G	0.88	0.12	0.0126	66018	1.92e-22	0.84	0.0664	0.0326	28104	0.04
rs10447226	5	139028810	G	T	0.75	-0.05	0.0092	73173	1.99e-8	0.72	-0.0208	0.0272	28104	0.45
rs10520313	4	176744539	G	A	0.92	-0.09	0.014	73173	8.73e-11	0.93	-0.0424	0.0477	28104	0.37
rs10736470	11	113418371	G	A	0.29	-0.07	0.009	73173	1.11e-14	0.36	0.0169	0.0256	28104	0.51
rs10774034	12	2330458	C	T	0.66	-0.08	0.0085	73173	7.10e-23	0.61	-0.054	0.0251	28104	0.03
rs10774545	12	110564659	A	C	0.28	0.06	0.0086	73173	6.98e-13	0.28	-0.0109	0.0274	28104	0.69
rs10861176	12	104631552	G	A	0.27	-0.05	0.0086	73173	5.23e-9	0.27	5,003-4	0.0280	28104	0.99
rs10861879	12	108609634	G	A	0.64	-0.05	0.0083	73173	1.23e-8	0.69	0.0397	0.0264	28104	0.13
rs10873538	14	104255569	T	G	0.66	-0.06	0.0082	73173	9.59e-15	0.66	0.0409	0.0258	28104	0.11
rs10894308	11	130891895	G	A	0.54	0.05	0.0079	73173	1.36e-9	0.55	-0.0263	0.0247	28104	0.29
rs10906984	15	83366296	A	C	0.39	0.05	0.0079	73173	4.84e-10	0.38	-0.0228	0.0255	28104	0.37
rs10985811	9	101070487	T	C	0.80	-0.05	0.0098	73173	2.53e-8	0.83	0.0114	0.0333	28104	0.73
rs11038906	11	46536442	G	A	0.85	-0.09	0.0108	73173	6.43e-16	0.83	-0.0047	0.0327	28104	0.89
rs11122119	1	6768856	C	A	0.65	-0.05	0.0081	73173	2.31e-8	0.68	-0.0419	0.0262	28104	0.11
rs11159633	14	84670453	T	G	0.37	-0.05	0.008	73173	1.89e-8	0.38	0.0589	0.0251	28104	0.02
rs11167584	5	152004727	A	C	0.78	0.05	0.0095	73173	5.47e-9	0.73	0.0045	0.0269	28104	0.87
rs11210203	1	73775513	A	G	0.36	0.06	0.0082	73173	2.31e-15	0.37	0.0146	0.0255	28104	0.57
rs1121296	1	72174197	T	C	0.59	0.05	0.008	73173	3.74e-9	0.63	-0.0245	0.0255	28104	0.34
rs112205523	19	50114350	C	T	0.71	0.06	0.0088	73173	1.82e-10	0.69	0.0613	0.0266	28104	0.02

rs1143702	1	44086831	C	T	0.33	0.07	0.0082	73173	2.76e-16	0.3296	-0.0012	0.0259	28104	0.9631
rs1157635	11	24404596	A	G	0.52	-0.04	0.0078	73173	2.00e-8	0.4917	0.05	0.0246	28104	0.04242
rs11696755	20	48105317	T	C	0.83	-0.07	0.0104	73173	2.99e-10	0.824	-0.0148	0.0318	28104	0.6422
rs11715134	3	50552866	G	A	0.89	-0.07	0.0123	73173	2.52e-28	0.9105	-0.0361	0.0418	28104	0.3878
rs11807834	1	230272624	G	A	0.76	-0.05	0.0093	73173	1.12e-8	0.7399	-0.0294	0.0281	28104	0.2962
rs11890137	2	198816662	G	A	0.36	0.06	0.0081	73173	7.34e-12	0.3253	-0.0883	0.0262	28104	7.61e-4
rs11899740	2	73631212	G	A	0.79	-0.06	0.0103	63155	9.94e-9	0.7783	-0.0209	0.0295	28104	0.479
rs1191551	14	30000405	T	G	0.18	0.06	0.0103	73173	4.72e-9	0.2229	-0.0089	0.0291	28104	0.7598
rs11956576	5	91001001	A	C	0.78	-0.05	0.0093	73173	1.00e-8	0.7968	-0.0334	0.0310	28104	0.2821
rs12126806	1	200963825	C	T	0.75	0.05	0.009	73173	2.89e-8	0.7062	0.0392	0.0267	28104	0.1434
rs12131135	1	200418613	G	T	0.55	0.04	0.0081	73173	3.68e-8	0.4924	-0.0274	0.0244	28104	0.2632
rs12244388	10	104640052	G	A	0.61	0.07	0.008	73173	1.59e-18	0.6684	4,00E-04	0.0279	28104	0.9886
rs12277680	11	134586708	A	G	0.51	-0.05	0.0081	73173.9	3.84e-11	0.5707	0.0132	0.0248	28104	0.5947
rs12301769	12	72231313	A	C	0.91	-0.08	0.014	72630.42	1.92e-9	0.90081	-0.0052	0.0418	28104	0.9011
rs12311848	12	124486851	A	G	0.70	-0.05	0.0087	73173.9	1.65e-8	0.6713	0.0486	0.0259	28104	0.06069
rs12614183	2	105005181	C	T	0.78	-0.05	0.0095	73173.9	4.17e-8	0.8232	-0.0128	0.0320	28104	0.6898
rs12652777	5	155775075	T	C	0.43	0.	0.0079	73173.9	1.06e-8	0.4916	0.0137	0.0246	28104	0.5781
rs12668848	7	2020995	G	A	0.56	0.08	0.0078	73173.9	3.21e-26	0.5709	0.0043	0.0247	28104	0.8618
rs12713008	2	48503561	G	A	0.55	0.04	0.0078	73173.9	3.05e-8	0.5428	0.001	0.0239	28104	0.9667
rs12883788	14	33303540	C	T	0.57	-0.05	0.0079	73173.9	8.43e-12	0.547	-0.0085	0.0244	28104	0.7281
rs12905223	15	85114268	T	C	0.75	0.067	0.009	73173.9	1.65e-10	0.7288	-0.0252	0.0273	28104	0.3569
rs12950148	17	12872802	A	G	0.83	-0.06	0.0105	73173.9	1.48e-09	0.8189	-0.0467	0.0316	28104	0.1406
rs12969453	18	52751708	A	G	0.56	0.05	0.0078	73173.9	3.53e-12	0.5989	0.0059	0.02509	28104	0.8141
rs13090130	3	161777035	G	A	0.62	0.05	0.0079	73173.9	9.92e-11	0.6343	-0.0143	0.02538	28104	0.5732
rs13107325	4	103188709	C	T	0.93	-0.16	0.0168	58439.19	2.9e-21	0.9263	0.0877	0.0482	28104	0.06889
rs13115045	4	143866615	G	A	0.78	0.05	0.0094	73173.9	4.36e-8	0.7626	0.0301	0.0288	28104	0.2976
rs13211901	6	27187612	G	A	0.84	0.11	0.011	71487.15	9.02e-26	0.798	-0.018	0.0299	28104	0.5476

rs13219424	6	128332084	C	T	0.67	0.05	0.0084	73173	4.52e-8	0.6986	-0.0287	0.0268	28104	0.285
rs13230189	7	137073259	C	T	0.668	0.07	0.0081	73173	1.04e-18	0.6427	0.0358	0.0258	28104	0.166
rs133293	22	42382797	G	A	0.45	-0.05	0.0081	73173	1.85e-11	0.5284	-0.0394	0.0246	28104	0.1106
rs1401988	12	103561799	C	T	0.67	0.058	0.0085	73173	4.63e-8	0.631	0.0145	0.0251	28104	0.564
rs1480380	6	32913246	C	T	0.92	0.15	0.0159	59322	4.75e-22	0.899	0.1107	0.0394	28104	5.07e-3
rs1553933	2	60509080	C	T	0.38	0.04	0.008	73173	3.89e-8	0.4104	-0.0162	0.0249	28104	0.5157
rs1600295	3	181234906	C	T	0.17	-0.07	0.0109	73173	1.29e-11	0.2058	-0.0061	0.0304	28104	0.8413
rs16825349	2	146425531	A	G	0.83	-0.07	0.0104	73173	1.32e-11	0.8161	0.0219	0.0314	28104	0.4856
rs169738	6	33537546	A	G	0.42	-0.05	0.0079	73173	4.39e-11	0.4003	-0.0234	0.0250	28104	0.3508
rs17194490	3	2547786	G	T	0.84	-0.08	0.0116	60303	1.85e-11	0.8454	-0.0316	0.0330	28104	0.3394
rs17693963	6	27710165	A	C	0.92	0.19	0.0151	64563	1.32e-37	0.8858	0.0283	0.0368	28104	0.4431
rs1784223	11	65380248	T	C	0.70	-0.05	0.0086	73173	2.43e-8	0.6536	0.0329	0.0257	28104	0.2017
rs1924378	13	38361744	T	G	0.48	0.05	0.0081	73173	9.31e-9	0.4139	-0.0011	0.0249	28104	0.9648
rs1939514	11	132572894	T	C	0.52	0.06	0.0077	73173	1.06e-12	0.4942	-0.0443	0.0246	28104	0.0721
rs209180	6	28793605	C	T	0.61	0.05	0.0083	73173	6.93e-10	0.6087	0.0618	0.0251	28104	0.01366
rs2095401	6	43124720	C	T	0.83	0.06	0.0106	73173	4.61e-8	0.8643	-0.0097	0.0361	28104	0.7884
rs2119242	10	21344773	G	A	0.83	-0.06	0.0106	73173	1.34e-8	0.8686	-0.012	0.0354	28104	0.7351
rs215483	4	23377121	G	A	0.69	-0.05	0.0084	73173	1.59e-9	0.6638	0.02	0.02620	28104	0.4453
rs2161203	5	88696876	T	C	0.89	0.08	0.0135	62066	2.49e-9	0.8772	0.0063	0.03664	28104	0.8635
rs216221	17	2155430	G	A	0.32	-0.05	0.0082	73173	8.65e-10	0.3494	-0.0409	0.0256	28104	0.1115
rs217291	6	84393942	T	C	0.57	0.05	0.0078	73173	5.47e-10	0.5632	-0.0072	0.02479	28104	0.7715
rs2190864	14	72416219	T	C	0.39	0.07	0.008	73173	1.12e-16	0.4004	-0.0316	0.0248	28104	0.2044
rs2224086	1	115309590	C	A	0.20	-0.06	0.0103	73173	2.09e-8	0.2276	-0.0215	0.02936	28104	0.464
rs2252074	7	104594253	T	G	0.59	-0.07	0.0078	73173	1.27e-14	0.6017	0.0252	0.0250	28104	0.314
rs2319280	1	150014017	A	C	0.15	-0.07	0.0118	61788	1.37e-8	0.1617	-0.0177	0.03266	28104	0.5879
rs2381411	9	36319928	T	C	0.57	-0.05	0.0079	73173.9	1.28e-8	0.6055	0.0055	0.0253	28104	0.8283
rs2455415	13	38860697	C	T	0.6	-0.05	0.0081	73173.9	3.431e-09	0.5697	0.0366	0.02490	28104	0.1417

rs246024	5	140333952	C	T	0.47	0.05	0.0077	73173.9	1.71e-09	0.4789	0.0513	0.02454	28104	0.03659
rs2470578	3	17330092	G	A	0.60	0.05	0.0085	73173.9	3.49e-08	0.5221	0.0114	0.0246	28104	0.644
rs2686386	12	121639657	C	T	0.21	0.05	0.0096	73173.9	1.26e-08	0.1837	0.0028	0.0320	28104	0.9303
rs2710323	3	52815905	T	C	0.53	0.07	0.0077	73173.9	5.92e-22	0.533	-0.0034	0.02483	28104	0.8911
rs2802535	1	98508258	C	T	0.17	-0.10	0.0105	73173.9	3.54e-20	0.1852	0.0628	0.0313	28104	0.04486
rs2815731	6	73155285	C	A	0.64	0.05	0.0081	73173.9	3.41e-10	0.6654	-0.0408	0.0259	28104	0.1153
rs281774	2	200814572	C	T	0.20	0.08	0.0096	73173.9	3.68e-17	0.1881	0.0772	0.0308	28104	0.01223
rs28669908	15	78910267	C	A	0.80	0.08	0.0098	73173.9	9.25e-15	0.7851	0.0296	0.02996	28104	0.3233
rs28758902	18	53408187	C	T	0.54	-0.05	0.0077	73173.9	7.69e-12	0.569	0.0239	0.02499	28104	0.339
rs28768122	12	123885974	T	C	0.26	0.08	0.009	73173.9	3.71e-18	0.2508	0.0511	0.02892	28104	0.07733
rs28865701	18	77620911	A	G	0.62	-0.06	0.0079	73173.9	4.7e-16	0.6344	-0.0274	0.02548	28104	0.2824
rs2909457	2	162845855	G	A	0.38	0.05	0.0083	73173.9	3.09e-8	0.4546	0.0599	0.02459	28104	0.01488
rs2910032	5	152540354	C	T	0.42	0.06	0.008	73173.9	4.52e-12	0.4826	0.0368	0.02456	28104	0.1341
rs2914983	2	198314568	A	G	0.35	0.06	0.0081	73173.9	1.1e-14	0.3244	-0.0706	0.0260	28104	6.78e-3
rs2968533	7	71787660	G	A	0.60	0.05	0.008	73173.9	4.593e-9	0.5627	-0.0411	0.02453	28104	0.09387
rs308697	3	161487491	C	A	0.60	0.056	0.0079	73173.9	3.35e-9	0.5715	-0.0306	0.02453	28104	0.2123
rs312477	3	53515136	G	A	0.746	0.06	0.0089	73173.9	1.37e-10	0.7758	-0.0624	0.02916	28104	0.03237
rs3129963	6	32380208	A	G	0.84	0.09	0.0113	70162.94	8.34e-15	0.8193	0.1026	0.0308	28104	8.97e-4
rs3131063	6	30763756	G	A	0.55	0.05	0.0082	69694.12	6.50e-9	0.5554	0.0617	0.0246	28104	0.01238
rs3134797	6	32186050	G	A	0.59	0.07	0.0083	70162.94	2.42e-11	0.5667	0.0499	0.02464	28104	0.0429
rs322128	19	11402416	C	T	0.78	-0.06	0.0095	73173.9	2.08e-09	0.8021	0.008	0.0309	28104	0.7962
rs337718	18	69774278	T	C	0.30	0.05	0.0084	73173.9	4.39e-09	0.2937	-0.035	0.02695	28104	0.1942
rs34829140	2	155904379	C	A	0.80	-0.05	0.0098	73173.9	4.04e-08	0.7834	0.0216	0.02947	28104	0.4637
rs35045093	7	127738471	A	C	0.826	0.06	0.0103	73173.9	3.36e-08	0.8248	-0.0379	0.03322	28104	0.254
rs35339313	2	145139727	C	T	0.61	-0.06	0.008	73173.9	1.31e-12	0.6544	0.0248	0.02558	28104	0.3324
rs353547	3	52268866	T	C	0.39	-0.06	0.0079	73173.9	5.17e-14	0.3695	-0.0028	0.02553	28104	0.9127
rs35531336	7	133438132	A	G	0.89	0.07	0.013	73173.9	1.46e-08	0.8765	-0.0251	0.0370	28104	0.4978

rs35734242	4	706700	T	C	0.57	-0.045	0.008	73173.9	3.927e-10	0.5683	9,00E-04	0.0251	28104	0.9714
rs35844949	2	58380167	G	A	0.44	-0.07	0.0078	73173.9	6.203e-21	0.4404	-0.0058	0.0249	28104	0.8162
rs36013966	8	89302448	G	A	0.78	-0.06	0.0093	73173.9	7.548e-10	0.7662	-0.0403	0.029	28104	0.1678
rs36109883	6	26133070	G	A	0.93	0.18	0.0174	58248.36	6.789e-24	0.91004	0.0362	0.0407	28104	0.3741
rs3732386	3	36871993	C	T	0.70	-0.06	0.0087	73173.9	1.464e-10	0.6418	0.0433	0.025	28104	0.09107
rs3739554	9	129940645	A	G	0.83	-0.06	0.0103	73173.9	2.257e-08	0.8385	-0.0034	0.033	28104	0.9194
rs3765971	1	8445360	C	T	0.28	0.05	0.0088	73173.9	1.975e-09	0.3382	-0.0131	0.025	28104	0.6132
rs3779870	8	65519166	A	G	0.37	-0.05	0.008	73173.9	7.63e-09	0.3731	0.0588	0.0255	28104	0.02144
rs3808581	8	26250047	G	A	0.79	-0.07	0.0097	73173.9	3.823e-12	0.8335	0.0564	0.033	28104	0.08788
rs3929747	10	104214267	C	T	0.75	0.06	0.0091	73173.9	3.127e-12	0.7905	0.0041	0.03071	28104	0.8938
rs39967	5	57744788	T	C	0.16	-0.06	0.0107	73173.9	1.869e-08	0.1771	-0.0376	0.03223	28104	0.2434
rs4144848	2	200001233	G	A	0.54	0.07	0.0079	73173.9	2.328e-16	0.5789	0.0307	0.0248	28104	0.2163
rs41533650	4	176869252	G	A	0.80	-0.07	0.0097	73173.9	8.691e-14	0.7906	-0.0196	0.0300	28104	0.5139
rs4283241	16	29988349	G	A	0.57	0.06	0.0078	73173.9	2.531e-14	0.541	-0.0886	0.0244	28104	2.91e-4
rs431982	5	106732144	T	C	0.24	0.05	0.0093	73173.9	1.558e-08	0.1817	0.0593	0.0322	28104	0.06609
rs4327046	16	9942377	G	A	0.73	-0.06	0.0088	73173.9	4.1e-10	0.7061	-0.0198	0.0274	28104	0.47
rs4510068	17	44184828	G	T	0.60	0.05	0.0082	73173.9	1.757e-09	0.608	-0.1634	0.0254	28104	1.28e-10
rs459391	21	22120508	T	C	0.19	0.06	0.01	73173.9	1.543e-08	0.1932	0.0408	0.0308	28104	0.1863
rs4712938	6	25449269	T	C	0.21	-0.06	0.0097	73173.9	4.446e-09	0.237	0.0031	0.02914	28104	0.9153
rs4793888	17	55737740	G	A	0.79	-0.06	0.0097	73173.9	3.634e-10	0.8478	-0.0307	0.03431	28104	0.371
rs4808955	19	19552413	T	C	0.66	-0.06	0.0081	73173.9	1.967e-13	0.646	0.0502	0.02578	28104	0.05159
rs4936215	11	133852684	A	G	0.81	0.08	0.0104	72178.92	1.857e-15	0.7972	0.006	0.0303	28104	0.8431
rs496108	5	49532628	C	T	0.45	0.05	0.0082	73173.9	8.617e-10	0.4659	0.0152	0.0247	28104	0.539
rs4975957	5	46265302	A	G	0.46	0.05	0.0084	72359.83	3.581e-09	0.4686	0.0162	0.02479	28104	0.5135
rs4990036	6	31323506	C	T	0.89	0.15	0.0135	64853.91	1.034e-27	0.8577	0.075	0.03397	28104	0.02726
rs500102	9	77358745	T	C	0.43	0.04	0.0079	73173.9	4.153e-08	0.3944	0.0151	0.0250	28104	0.5467
rs505061	9	22767164	C	A	0.51	-0.05	0.0077	73173.9	1.025e-10	0.4962	-0.0162	0.02430	28104	0.5051

rs55704727	3	117616431	C	T	0.90	0.07	0.013	72630.42	3.296e-08	0.8757	0.0246	0.0365	28104	0.5014
rs56086712	10	3800240	T	C	0.50	0.04	0.0081	73173.9	4.389e-08	0.4377	0.0199	0.0251	28104	0.4294
rs5757730	22	39967430	A	G	0.39	-0.05	0.0085	73173.9	1.185e-10	0.4986	0.0427	0.0245	28104	0.08162
rs578470	12	50463325	T	C	0.62	-0.05	0.0083	73173.9	2.327e-08	0.5828	-0.0189	0.0248	28104	0.4467
rs58946679	1	2372321	C	T	0.61	-0.05	0.0084	73173.9	1.721e-10	0.6475	0.0067	0.0262	28104	0.7985
rs6010045	22	51103091	T	C	0.32	-0.05	0.0085	72992.88	1.815e-08	0.3176	0.0192	0.0264	28104	0.4681
rs6065094	20	37453194	A	G	0.32	-0.06	0.0082	73173.9	1.409e-14	0.3203	0.0074	0.0259	28104	0.7751
rs61405217	4	170322312	C	T	0.42	0.05	0.0079	73173.9	5.387e-11	0.4588	0.0035	0.0247	28104	0.8876
rs61789528	1	97157335	G	A	0.81	-0.06	0.0098	73173.9	8.782e-09	0.8214	-0.0191	0.0325	28104	0.5573
rs61828917	1	173580303	C	T	0.85	0.07	0.011	73173.9	7.951e-10	0.8812	0.0594	0.0374	28104	0.1125
rs61896127	11	47099867	C	T	0.91	-0.10	0.0151	59339.13	1.238e-10	0.91695	0.0297	0.04353	28104	0.4951
rs61937595	12	57682956	C	T	0.92	0.12	0.0158	69922.48	1.317e-14	0.9149	0.0096	0.0436	28104	0.8259
rs62176172	2	185499482	C	T	0.64	0.05	0.0081	73173.9	5.03e-10	0.6814	0.0152	0.02636	28104	0.5642
rs62184960	2	172974789	C	T	0.88	0.07	0.0122	73173.9	1.082e-08	0.8459	0.0027	0.0337	28104	0.9363
rs634940	6	93077500	G	T	0.79	-0.06	0.0098	73173.9	2.884e-11	0.7453	-0.0659	0.0278	28104	0.018
rs6482437	10	18726326	A	C	0.09	-0.10	0.0135	73173.9	1.05e-14	0.103	0.0296	0.03876	28104	0.4451
rs6500602	16	4497451	T	C	0.30	0.05	0.0084	73173.9	4.769e-08	0.2975	0.012	0.02699	28104	0.6567
rs6504163	17	61545779	C	T	0.37	-0.05	0.0082	73173.9	1.869e-09	0.3626	0.0028	0.02545	28104	0.9124
rs6538539	12	95195293	G	T	0.47	0.05	0.0077	73173.9	5.634e-10	0.4501	0.001	0.0252	28104	0.9684
rs6588168	1	66324118	C	T	0.48	-0.059	0.0079	73173.9	5.941e-10	0.4433	-0.0096	0.02488	28104	0.6997
rs6670165	1	177280121	C	T	0.81	-0.07	0.0098	73173.9	1.078e-11	0.81	-0.0547	0.03119	28104	0.07949
rs668437	3	135941739	G	A	0.42	0.05	0.0081	73173.9	9.253e-10	0.4735	-0.0461	0.0247	28104	0.06267
rs6713590	2	193937515	C	T	0.47	0.04	0.0077	73173.9	2.694e-08	0.481	-0.0335	0.0245	28104	0.1721
rs6715366	2	2327295	G	A	0.77	-0.06	0.0094	73173.9	4.225e-09	0.7258	0.0401	0.02728	28104	0.1416
rs67627854	2	76246645	T	G	0.82	-0.06	0.0101	73173.9	2.202e-08	0.7936	0.0167	0.0303	28104	0.5821
rs6839635	4	103872854	C	A	0.50	-0.04	0.0079	73173.9	3.874e-08	0.5082	0.0606	0.0245	28104	0.01339
rs6848123	4	80203425	A	C	0.56	0.05	0.0081	73173.9	2.054e-09	0.4971	-0.0035	0.02454	28104	0.8866

rs6919146	6	165059966	T	G	0.31	-0.05	0.0085	73173.9	8.423e-09	0.3404	-0.0242	0.02601	28104	0.3522
rs6925079	6	64946311	T	C	0.64	-0.04	0.0081	73173.9	3.581e-08	0.6802	-0.0154	0.02621	28104	0.5569
rs6926151	6	96476028	A	G	0.94	0.10	0.0176	59292.61	3.327e-08	0.92955	0.0647	0.04989	28104	0.1947
rs6943762	7	86403263	T	C	0.87	0.10	0.0124	73173.9	6.296e-17	0.8791	-0.0507	0.03707	28104	0.1715
rs6969410	7	110069015	T	G	0.66	0.06	0.0083	73173.9	1.899e-11	0.6211	-8.00e-04	0.02492	28104	0.9744
rs6984242	8	60700469	G	A	0.40	0.05	0.0078	73173.9	1.5e-11	0.4009	0.0127	0.0249	28104	0.6103
rs7002992	8	103676605	T	C	0.67	0.05	0.0083	73173.9	4.483e-09	0.6297	0.0116	0.02540	28104	0.6479
rs7007361	8	111621265	C	T	0.84	-0.07	0.0104	73173.9	4.429e-11	0.8274	-0.0144	0.0323	28104	0.6564
rs708228	11	57585662	C	T	0.70	-0.05	0.0085	73173.9	2.062e-09	0.6864	-0.0129	0.02617	28104	0.6221
rs7112912	11	134246900	C	T	0.41	0.05	0.0084	73173.9	2.09e-10	0.4674	-0.0502	0.02473	28104	0.04241
rs7113068	11	124633529	T	C	0.86	0.08	0.0112	73173.9	4.605e-12	0.8367	0.0313	0.03369	28104	0.3529
rs715170	18	53795514	C	T	0.75	0.06	0.009	73173.9	7.4e-13	0.743	-0.0392	0.0277	28104	0.1584
rs7177338	15	91428636	G	A	0.42	-0.07	0.0082	73173.9	2.024e-19	0.4703	-8.00e-04	0.0243	28104	0.9738
rs7178152	15	89787055	T	G	0.53	0.05	0.0078	73173.9	2.628e-10	0.5362	-0.0302	0.0244	28104	0.2177
rs717947	4	33653757	C	T	0.87	-0.07	0.0116	73173.9	3.096e-09	0.90095	-0.0537	0.0430	28104	0.2121
rs7207904	17	1267857	G	A	0.61	0.05	0.0079	73173.9	1.645e-09	0.6194	0.0155	0.0252	28104	0.5394
rs7238071	18	77579812	A	G	0.69	-0.0	0.0084	73173.9	9.292e-14	0.7263	-0.0452	0.02797	28104	0.1062
rs72728416	1	97834691	A	G	0.73	-0.06	0.0087	73173.9	4.99e-12	0.7566	-0.0042	0.0285	28104	0.8829
rs72761691	9	134786548	A	C	0.86	-0.07	0.0116	73173.9	1.07e-08	0.8598	0.0053	0.03540	28104	0.881
rs72769124	1	239210058	C	A	0.91	-0.10	0.0143	73173.9	4.36e-13	0.90654	0.0167	0.04214	28104	0.6919
rs72974269	2	225454907	C	T	0.68	0.05	0.0083	73173.9	2.76e-10	0.695	-0.0494	0.02648	28104	0.06219
rs7312697	12	29933069	T	C	0.36	-0.05	0.0081	73173.9	4.85e-11	0.3745	-0.0139	0.0253	28104	0.5838
rs73229090	8	27442127	C	A	0.89	0.10	0.0142	58749.13	4.34e-13	0.8933	0.0578	0.03875	28104	0.1359
rs74695911	7	131541469	T	C	0.82	0.06	0.0103	73173.9	4.897e-08	0.784	0.0373	0.02998	28104	0.2135
rs7563610	2	37566376	A	G	0.58	0.05	0.0079	73173.9	4.48e-10	0.6052	-0.0307	0.02504	28104	0.2202
rs758749	19	57189718	C	T	0.86	-0.06	0.0112	73173.9	4.66e-8	0.8696	0.0116	0.03733	28104	0.756
rs7600146	2	22543347	G	T	0.63	0.05	0.0083	73173.9	8.311e-09	0.569	-0.013	0.02485	28104	0.601

rs7609876	3	176790116	T	C	0.72	-0.05	0.0089	73173.9	9.398e-9	0.7932	0.0313	0.03031	28104	0.3018
rs7632834	3	17886678	T	C	0.42	-0.05	0.0085	73173.9	2.091e-9	0.5062	-0.0099	0.02461	28104	0.6875
rs7647398	3	180733150	C	T	0.82	0.09	0.01	73173.9	2.207e-17	0.8063	0.0122	0.03067	28104	0.6908
rs7703010	5	137711716	C	A	0.37	0.05	0.0082	73173.9	8.278e-09	0.4127	-0.0028	0.0247	28104	0.9101
rs7703618	5	44914579	G	A	0.44	0.05	0.008	73173.9	3.408e-09	0.3793	-0.0051	0.02550	28104	0.8415
rs7706116	5	45742624	G	T	0.56	-0.06	0.0078	73173.9	6.786e-13	0.5612	-0.0115	0.02502	28104	0.6458
rs77154381	2	27982119	A	G	0.68	0.05	0.0086	73173.9	2.6e-08	0.7468	0.0184	0.02833	28104	0.5161
rs778371	2	233743109	A	G	0.746	-0.05	0.0089	73173.9	1.103e-16	0.7248	0.016	0.02743	28104	0.5597
rs7811417	7	21534152	T	C	0.35	0.05	0.0081	73173.9	2.166e-09	0.33	-8,00E-04	0.02681	28104	0.9762
rs7816998	8	38257506	G	A	0.77	0.06	0.0092	73173.9	3.105e-10	0.7789	0.019	0.02893	28104	0.5114
rs7915131	10	64418656	C	T	0.42	0.04	0.0078	73173.9	4.943e-08	0.4319	0.051	0.02469	28104	0.03894
rs79210963	7	24717969	T	C	0.90	-0.09	0.0129	73173.9	2.584e-11	0.8922	-0.0214	0.03880	28104	0.5813
rs7927176	11	123395864	A	G	0.66	-0.05	0.0081	73173.9	6.604e-11	0.6617	0.0139	0.02606	28104	0.5938
rs8055219	16	13753384	G	A	0.79	-0.07	0.0095	73173.9	1.565e-12	0.7785	-0.0525	0.0295	28104	0.07584
rs8104557	19	31030189	T	C	0.82	-0.06	0.011	61674.31	3.764e-08	0.8023	-0.0594	0.03067	28104	0.05282
rs8180995	8	143326237	A	G	0.53	0.06	0.0078	73173.9	2.19e-15	0.5216	0.0405	0.0246	28104	0.1005
rs832195	3	63864612	T	C	0.34	0.05	0.0083	73173.9	2.969e-09	0.2972	0.0593	0.0271	28104	0.02872
rs9257802	6	29343355	C	T	0.83	0.09	0.0107	70162.94	1.182e-15	0.8034	0.0493	0.0306	28104	0.108
rs9287971	2	174931752	G	A	0.60	-0.05	0.0083	73173.9	3.824e-08	0.671	0.0407	0.026	28104	0.131
rs9304548	18	27500959	C	A	0.26	0.06	0.0089	73173.9	1.895e-11	0.2623	0.0041	0.0278	28104	0.8829
rs9318638	13	79986153	T	C	0.43	-0.05	0.0079	73173.9	2.044e-09	0.477	0.0169	0.0245	28104	0.4914
rs9393741	6	26654189	A	G	0.82	0.09	0.0102	73173.9	5.897e-19	0.7878	-0.0024	0.029	28104	0.9351
rs9454727	6	70003389	A	G	0.77	0.05	0.0098	70865.66	1.93e-08	0.7358	0.0164	0.02774	28104	0.5544
rs9569820	13	58702746	G	T	0.84	-0.07	0.0109	73173.9	6.563e-10	0.8836	0.0254	0.03889	28104	0.5137
rs9597388	13	56928696	G	A	0.82	0.07	0.0101	73173.9	3.239e-11	0.8067	0.0044	0.03113	28104	0.8876
rs9623320	22	41478482	G	A	0.76	-0.06	0.0092	73173.9	2.371e-10	0.7122	-0.045	0.0270	28104	0.09665
rs967005	6	28210688	C	T	0.83	0.07	0.0102	73173.9	1.3e-11	0.8046	0.0183	0.03081	28104	0.5526

rs9803993	1	243672125	G	A	0.78	0.05	0.0093	73173.9	1.497e-08	0.7811	0.0242	0.02959	28104	0.4136
rs9831201	3	80913228	T	G	0.72	0.054	0.0087	73173.9	6.7e-10	0.6993	-0.0132	0.02657	28104	0.6194
rs9882532	3	16865845	T	C	0.6	-0.0	0.0087	73173.9	8.574e-10	0.6518	-0.0494	0.02587	28104	0.05623
rs9891739	17	19942177	C	T	0.54	-0.04	0.008	73173.9	2.181e-08	0.4702	-0.0815	0.02449	28104	0.0008751
rs9926296	16	89818089	A	G	0.44	0.05	0.0086	71743.28	2.374e-08	0.49	0.0386	0.02457	28104	0.1162
rs9975024	21	16439883	A	G	0.53	-0.05	0.008	73173.9	1.777e-09	0.5254	0.012	0.02445	28104	0.6237
rs999494	2	73157395	C	T	0.82	0.06	0.0101	73173.9	1.044e-08	0.7975	-0.0032	0.03070	28104	0.917

Instrumental SNPs select for the SCZ to BAG classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S12. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the BIP to BAG MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10043984	5	137712121	C	T	0.76	-0.06	0.0108	50981.48	3.709e-08	0.7468	-0.0202	0.0277	28104	0.467
rs10994299	10	62076628	C	A	0.95	-0.13	0.0201	50862.93	9.587e-11	0.94999	0.0062	0.0577	28104	0.9145
rs113779084	7	11871787	G	A	0.70	-0.08	0.0102	50981.48	1.423e-13	0.7099	-0.0301	0.0273	28104	0.2716
rs11647445	16	9926966	T	G	0.67	-0.05	0.0098	50981.48	2.152e-08	0.6472	-0.0169	0.0257	28104	0.5109
rs11764361	7	105043229	A	G	0.67	0.06	0.0104	50981.48	3.471e-09	0.6565	-0.0357	0.0258	28104	0.1674
rs11856299	15	83534421	C	T	0.75	0.06	0.011	50981.48	1.528e-08	0.7649	-0.0026	0.0295	28104	0.9298
rs12289486	11	79092527	C	T	0.89	-0.08	0.0149	50981.48	3.297e-08	0.892	0.0176	0.0389	28104	0.6517
rs1235162	6	29537224	A	G	0.91	0.11	0.0164	49000.67	6.899e-12	0.8804	0.0822	0.03623	28104	0.02328
rs12575685	11	70517927	G	A	0.67	-0.07	0.0101	50981.48	1.243e-10	0.7012	-0.0301	0.02643	28104	0.2549
rs12668848	7	2020995	G	A	0.58	0.06	0.0095	50981.48	1.902e-09	0.5709	0.0043	0.0247	28104	0.8618
rs12672003	7	24647222	A	G	0.89	-0.09	0.0155	46163.71	2.718e-09	0.8924	-0.0139	0.03892	28104	0.721
rs13044225	20	60865815	A	G	0.56	-0.05	0.0095	50186.29	8.498e-09	0.5319	0.0124	0.0244	28104	0.6127
rs13195401	6	26463574	G	T	0.92	0.14	0.0175	50615.04	6.979e-15	0.90102	0.039	0.0389	28104	0.3163
rs1487445	6	98565211	C	T	0.51	-0.07	0.0093	50981.48	1.484e-15	0.5324	0.0098	0.02467	28104	0.6913
rs17183814	2	166152389	G	A	0.92	0.10	0.0185	50772.41	2.678e-08	0.92605	0.0694	0.0464	28104	0.1354
rs17202645	2	169399768	C	T	0.81	0.07	0.0121	50981.48	2.494e-08	0.8177	-0.0273	0.0319	28104	0.3923
rs174581	11	61606683	G	A	0.63	-0.07	0.0097	50981.48	4.944e-13	0.6495	-0.0509	0.0256	28104	0.04681
rs17526722	6	25918855	G	A	0.93	0.10	0.0181	50615.04	1.759e-08	0.91019	0.0409	0.04075	28104	0.3156
rs17693963	6	27710165	A	C	0.91	0.12	0.0161	50772.41	5.836e-14	0.8858	0.0283	0.036	28104	0.4431
rs1894401	15	91429042	G	A	0.46	-0.05	0.0096	50035.72	2.799e-08	0.4709	1,00E-04	0.01899	28104	0.9958
rs1998821	10	18751242	G	A	0.89	0.08	0.0152	50981.48	4.25e-08	0.8948	-0.0299	0.038	28104	0.4395
rs2019611	15	85122620	C	A	0.75	0.07	0.0108	50981.48	1.437e-10	0.7509	-0.0423	0.0282	28104	0.1338
rs2069772	4	123373133	T	C	0.73	-0.06	0.0103	50981.48	1.567e-08	0.7318	-0.0386	0.0275	28104	0.1616
rs2126180	1	61105668	G	A	0.54	-0.06	0.0094	50981.48	1.621e-09	0.5298	-3,00E-04	0.0232	28104	0.9897
rs2159100	12	2346393	C	T	0.67	-0.08	0.0098	50981.48	1.872e-15	0.6682	-0.04	0.02614	28104	0.126

rs237475	20	48049506	T	C	0.52	-0.05	0.0093	50981.48	5.171e-09	0.5201	0.0134	0.02457	28104	0.5856
rs2577831	3	52628056	C	A	0.51	0.07	0.0093	50981.48	3.834e-13	0.5291	-0.003	0.02414	28104	0.9011
rs2719164	2	194437889	A	G	0.57	0.05	0.0095	50981.48	4.847e-08	0.5763	-0.011	0.02488	28104	0.6585
rs28565152	5	7542911	G	A	0.767	-0.07	0.0112	48268.74	1.963e-09	0.7557	-0.0354	0.02874	28104	0.2181
rs2899059	15	42844361	C	T	0.927	-0.1	0.017	50981.48	1.243e-08	0.91467	0.0753	0.04554	28104	0.09826
rs2953928	8	34152492	G	A	0.937	-0.12	0.02	50615.04	6.249e-09	0.94541	-0.0462	0.05104	28104	0.3654
rs3094035	6	30363136	G	T	0.907	0 1	0.0153	49000.67	4.168e-10	0.8645	0.0588	0.03447	28104	0.0881
rs3115674	6	31799076	T	G	0.897	0.08	0.0154	48295.99	3.871e-08	0.8713	0.0882	0.03535	28104	0.01261
rs3130557	6	31094703	C	T	0.89	0.09	0.0147	50862.93	9.734e-11	0.8564	0.074	0.033624	28104	0.02775
rs35741362	6	27007687	T	C	0.89	0.10	0.0148	50862.93	1.366e-12	0.8638	0.0135	0.0347	28104	0.6978
rs35958438	15	38973793	G	A	0.77	0.0	0.0117	47077.45	3.83e-08	0.7742	0.0192	0.0293	28104	0.5131
rs3856207	1	163747095	A	G	0.45	0.05	0.0095	50186.29	1.156e-08	0.4603	-0.0053	0.0244	28104	0.8284
rs39824	5	169255522	G	A	0.33	-0.06	0.0099	50981.48	9.568e-09	0.343	0.0217	0.0258	28104	0.4018
rs4072458	3	36862680	T	C	0.46	-0.08	0.0094	50981.48	5.233e-17	0.4529	0.0167	0.0246	28104	0.4973
rs41315395	6	32201469	C	A	0.86	-0.07	0.0127	49209.73	1.476e-08	0.8502	0.0314	0.03501	28104	0.3698
rs4619651	2	97416153	G	A	0.67	0.07	0.0101	50981.48	4.776e-11	0.6917	-0.0239	0.02633	28104	0.3642
rs4672	11	64009879	G	A	0.92	-0.10	0.0172	50035.72	3.42e-09	0.92068	-0.1075	0.04518	28104	0.01735
rs4790841	17	1835482	C	T	0.85	-0.07	0.0132	50035.72	3.141e-08	0.8462	0.0367	0.03379	28104	0.2775
rs531490	11	66324360	T	G	0.46	0.05	0.0093	50981.48	6.016e-09	0.456	-0.0145	0.02441	28104	0.5526
rs5758064	22	41153879	T	C	0.52	0.05	0.0093	50981.48	2.011e-08	0.5221	0.029	0.02474	28104	0.2413
rs61554907	17	38220432	G	T	0.88	-0.09	0.0154	50035.72	1.636e-08	0.8947	-0.0842	0.0395	28104	0.03331
rs67712855	20	43682551	T	G	0.69	0.07	0.0103	50186.29	4.221e-11	0.7129	0.0039	0.02710	28104	0.8856
rs6806239	3	70488207	T	G	0.80	0.07	0.0119	50981.48	2.641e-08	0.8163	-0.0315	0.03145	28104	0.3166
rs6865469	5	78849505	G	T	0.72	-0.06	0.0103	50981.48	1.65e-08	0.7309	-0.0366	0.02798	28104	0.1909
rs6887473	5	80961069	G	A	0.74	0.06	0.0105	50981.48	8.814e-09	0.7178	-0.0204	0.0272	28104	0.4546
rs6954854	7	21492589	G	A	0.43	0.06	0.0094	50981.48	5.936e-10	0.426	-0.0075	0.0249	28104	0.7634
rs7461753	8	144981498	A	C	0.6	-0.06	0.0098	50981.48	5.24e-09	0.5842	-0.0182	0.02501	28104	0.4668

rs763193	6	166996127	A	G	0.51	-0.0	0.0094	50981.48	1.192e-08	0.5185	-0.0142	0.0248	28104	0.5676
rs7813581	8	9882652	G	T	0.83	-0.08	0.0119	50981.48	9.911e-11	0.8286	0.0275	0.0325	28104	0.3988
rs9393926	6	28682725	G	A	0.09	-0.11	0.0165	49000.67	7.501e-12	0.1176	-0.031	0.0367	28104	0.3987

Instrumental SNPs select for the BIP to BAG classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S13. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the MDD to BAG MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs12658032	5	103904226	A	G	0.36	0.05	0.0083	69115.85	1.181e-10	0.3497	0.0225	0.0257	28104	0.3804
rs1460943	1	72813129	T	C	0.61	0.04	0.0081	69115.85	4.27e-8	0.609	-0.0145	0.0249	28104	0.5611
rs148696809	6	28934352	T	C	0.90	0.08	0.0132	69115.85	4.355e-9	0.8902	0.0412	0.037	28104	0.2712
rs2799077	6	28234597	T	C	0.15	-0.07	0.0112	69115.85	2.864e-9	0.1637	-0.0213	0.0327	28104	0.5153
rs4593766	1	73773043	T	C	0.38	0.05	0.0081	69115.85	1.705e-8	0.3694	0.0147	0.0254	28104	0.563
rs58825580	6	26365679	T	G	0.88	0.07	0.012	69115.85	1.258e-8	0.8664	0.0422	0.0349	28104	0.2272
rs6934329	6	27158033	A	G	0.82	0.06	0.0104	69115.85	1.416e-9	0.8128	-0.008	0.0308	28104	0.7952
rs6940116	6	27708732	A	G	0.84	0.06	0.0107	69115.85	4.912e-8	0.8228	-0.0048	0.0316	28104	0.8795
rs8013655	14	42089908	T	C	0.47	0.04	0.0079	69115.85	2.672e-8	0.4774	0.0518	0.0247	28104	0.03578

Instrumental SNPs select for the MDD to BAG classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S14. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the AD to BAG MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs10412413	19	45327309	T	C	0.3	0.0526	0.0023	429332.77	6.97e-115	0.32	0.0203	0.0264	28104	0.44
rs10808026	7	143099133	A	C	0.202	-0.0161	0.0027	425571.98	2.82e-9	0.22	-0.0219	0.03	28104	0.47
rs11257240	10	11719074	G	T	0.3	0.0129	0.0023	423921.06	1.28e-8	0.36	-0.0021	0.0252	28104	0.93
rs113260531	17	5138980	A	G	0.13	0.02	0.0033	429194.03	7.91e-10	0.12	0.0018	0.0376	28104	0.96
rs11672748	19	45490192	G	A	0.34=	0.023	0.0023	427777.91	7.25e-24	0.32	-0.026	0.0263	28104	0.32
rs117618017	15	63569902	T	C	0.12=	0.018	0.0033	428753.23	3.44e-8	0.14	0.0015	0.0345	28104	0.97
rs12539172	7	100091795	T	C	0.31=	-0.0167	0.0023	429961.08	5.85e-13	0.32	-0.0605	0.0261	28104	2.06e-2
rs12590654	14	92938855	A	G	0.34	-0.0148	0.0023	420451.85	1.32e-10	0.34	-0.0367	0.0263	28104	0.16
rs1532278	8	27466315	T	C	0.38	-0.0201	0.0022	428741.91	2.07e-19	0.39	0.0092	0.0251	28104	0.714
rs2081545	11	59958380	A	C	0.3	-0.0179	0.0022	425881.39	1.11e-15	0.40	-0.0624	0.025	28104	1.23e-2
rs2647066	6	32571122	T	C	0.13	-0.0208	0.0032	428954.21	8.55e-11	0.18	0.0101	0.0314	28104	0.75
rs28394864	17	47450775	A	G	0.45	0.0123	0.0022	424326.36	1.68e-8	0.45	0.0749	0.0248	28104	2.50e-3
rs28469095	19	45655333	C	T	0.1	0.0466	0.0036	426787.79	1.07e-38	0.09	0.0769	0.0427	28104	7.17e-2
rs28834970	8	27195121	C	T	0.37	0.0153	0.0022	428248.62	8.49e-12	0.36	0.0151	0.0254	28104	0.55
rs3752231	19	1043638	T	C	0.25	0.0159	0.0025	424712.83	2.38e-10	0.25	0.0665	0.0284	28104	1.93e-2
rs3810140	19	45316807	T	C	0.07	-0.0286	0.0043	424090.94	4.54e-11	0.06	0.0105	0.0562	28104	0.85
rs412776	19	45379516	A	G	0.11	0.02834	0.0035	428640.06	8.10e-16	0.12	-0.023	0.0379	28104	0.54
rs4575098	1	161155392	A	G	0.23	0.01641	0.0026	427808.62	1.90e-10	0.23	0.0763	0.0291	28104	8.69e-3
rs4663105	2	127891427	C	A	0.43	0.0311	0.0022	418699.97	1.45e-44	0.42	-0.0275	0.0252	28104	0.28
rs4803750	19	45247627	G	A	0.06	-0.0491	0.0044	427688.71	1.61e-28	0.07	0.0258	0.0481	28104	0.59
rs59007384	19	45396665	T	G	0.2	0.1389	0.0027	427829.43	1,00e-200	0.21	0.103	0.0306	28104	7.57e-4
rs593742	15	59045774	G	A	0.31	-0.0138	0.0023	426926.64	3.73e-9	0.31	0.0281	0.0267	28104	0.29
rs6014724	20	54998544	G	A	0.1	-0.0229	0.0037	427344.83	5.38e-10	0.09	0.0096	0.0433	28104	0.82
rs6448453	4	11026028	A	G	0.26	0.0147	0.0025	429961.08	1.98e-9	0.26	0.032	0.0276	28104	0.25
rs679515	1	207750568	T	C	0.17	0.0254	0.0027	429226.74	6.83e-19	0.17	0.0781	0.032	28104	1.47e-2

rs74862042	19	45217784	C	A	0.161	0.0281	0.0029	429642.96	9.11e-22	0.15	0.0142	0.0342	28104	0.68
rs846881	19	45078553	C	A	0.21	-0.0174	0.0027	422334.58	9.89e-11	0.21	-0.0278	0.0301	28104	0.36
rs867611	11	85776544	G	A	0.32	-0.0204	0.0023	427263.57	1.48e-18	0.31	0.0186	0.0263	28104	0.48
rs9381563	6	47432637	C	T	0.36	0.0145	0.0023	422732.93	1.99e-10	0.37	0.0211	0.0257	28104	0.41
rs9653111	19	45592475	T	C	0.11	-0.0194	0.0034	427366.64	1.22e-8	0.12	0.0228	0.0364	28104	0.53

Instrumental SNPs select for the AD to BAG classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S15. Harmonized instrumental SNPs ($p<5x10^{-8}$) for the PD to BAG MR.

SNP	CHR	POS	A1	A2	A1_exp_frq	Beta_exp	SE_exp	N_exp	P_exp	A1_out_frq	Beta_out	SE_out	N_out	P_out
rs105137893		182760073	T	G	0.82	0.1596	0.0219	62649	3.185e-13	0.80	0.0217	0.0312	28104	0.4874
rs119429114		90601310	A	G	0.92	-0.1872	0.0331	62649	1.566e-08	0.92	0.0745	0.0474	28104	0.1162
rs12817488	12	123296294	A	G	0.46	0.1032	0.0178	62649	7.111e-09	0.46	0.0452	0.0248	28104	0.06877
rs1372518	4	90757294	A	C	0.21	-0.2334	0.0209	62649	5.354e-29	0.22	0.0511	0.0299	28104	0.08718
rs1474055	2	169110394	T	C	0.13	0.1763	0.0248	62649	1.144e-12	0.12	-0.0547	0.0385	28104	0.1559
rs329652	11	133769699	A	G	0.36	0.1082	0.0174	62649.	5.027e-10	0.35	-0.0014	0.0256	28104	0.9565
rs343118664		951947	T	C	0.80	-0.2272	0.0231	62649	7.974e-23	0.83	-0.0409	0.0324	28104	0.207
rs351187626		32560631	T	C	0.16	-0.2042	0.0312	60613	6.026e-11	0.18	0.0123	0.0312	28104	0.6943
rs356175	4	90630814	T	C	0.66	-0.1538	0.0181	62649	1.607e-17	0.68	-0.0033	0.0261	28104	0.8997
rs4566208	17	16010920	A	G	0.43	0.0957	0.0174	62649	3.884e-08	0.44	0.0318	0.0246	28104	0.1967
rs4588066	18	40672964	A	G	0.33	0.1046	0.0178	62649	4.453e-09	0.32	0.0147	0.0264	28104	0.5781
rs4631042	4	15712550	A	C	0.53	0.1213	0.0169	62649	6.64e-13	0.52	0.0055	0.0249	28104	0.8254
rs4774417	15	61993702	A	G	0.74	0.1052	0.0192	62649	4.626e-08	0.73	-0.0355	0.0282	28104	0.2081
rs562369141	7	43483551	T	C	0.19	-0.2147	0.0258	62649	8.739e-17	0.17	0.2877	0.0316	28104	9.633e-20
rs620490	8	16697579	T	G	0.72	0.1174	0.019	62649	6.456e-10	0.752	0.0072	0.0270	28104	0.79
rs7210219	17	44018519	T	C	0.78	0.2607	0.0275	60613	2.657e-21	0.79	-0.2681	0.0301	28104	5.582e-19
rs763396985		60372588	A	G	0.89	-0.1818	0.0263	62649	5.00e-12	0.90	0.0423	0.0422	28104	0.3171
rs76904798	12	40614434	T	C	0.15	0.1352	0.0235	62649	9.224e-09	0.15	-0.0563	0.0348	28104	0.1059
rs7695720	4	77183300	A	C	0.79	0.1255	0.0208	62649	1.528e-09	0.78	0.0567	0.0294	28104	0.05426
rs858290	7	23248036	A	G	0.61	0.1036	0.0177	62649	4.443e-09	0.60	0.0166	0.0251	28104	0.5095

Instrumental SNPs select for the PD to BAG classical Mendelian randomization analysis. CHR: chromosome; POS: hg19 genomic position; A1: effective allele; A2: the other allele; A1_exp_frq: A1 frequency in the exposure dataset; Beta_exp, effect of A1 for the exposure; SE_exp; standard error for Beta_exp; N_exp: sample size for exposure; P_exp, association p value for exposure; A1_out_frq: A1 frequency in the outcome dataset; Beta_out: effect size of A1 on outcome; SE_out: standard error of Beta_out; N_out: sample size for outcome; P_out: association p value for outcome.

Table S16. Statistical finemapping for associated loci.

Locus	Lead SNP	PP_1	PP_2	PPI_1	SNP2	PPI_2
Chr3:183892867-183975709	rs73185796	0.51	0.46	0.7	rs11402585	0.07
Chr4:38591172-38779512	rs13132853	0.74	0.26	0.16	rs11727467	0.11
Chr5:78388694-78451813	rs79107704	0.75	0.25	0.48	rs185486757	0.03
Chr6:45407654-45511945	rs2790102	0.84	0.16	0.05	rs1934327	0.04
Chr8:124661974-124682971	rs7461069	0.82	0.18	0.1	rs13340533	0.09
Chr10:134544247-134597265	rs4880424	0.85	0.15	0.08	rs12767391	0.04
Chr14: 88391116-88556525	rs17203398	0.5	0.5	0.04	rs10137195	0.04
Chr17: 43101281-44863413	rs2106786	0.85	0.15	0.01	rs62062797	0.01

PP_1: posterior probability for the causal structural assuming only one causal variant in the locus. PP_2: posterior probability for the causal structural assuming only two causal variants. PPI_1: posterior probability of inclusion to the one-causal variant structure; PPI_2: posterior probability of inclusion for the ranked second SNP to the one-causal variant structure.

Table S17. eQTL statistics for rs73185796 from GTEx v8 portal.

Gencode Id	Gene Symbol	Variant Id	P-Value	NES	Tissue
ENSG00000145198.14	VWA5B2	chr3_184257921_G_T_b38	8.4e-6	-0.67	Brain - Cortex
ENSG00000161202.17	DVL3	chr3_184257921_G_T_b38	1.2e-5	-0.23	Thyroid
ENSG00000145198.14	VWA5B2	chr3_184257921_G_T_b38	1.9e-5	-0.57	Nerve - Tibial
ENSG00000145198.14	VWA5B2	chr3_184257921_G_T_b38	2.2e-5	-0.74	Brain - Frontal Cortex (BA9)
ENSG00000161202.17	DVL3	chr3_184257921_G_T_b38	2.9e-5	-0.23	Cells - Cultured fibroblasts
ENSG00000161202.17	DVL3	chr3_184257921_G_T_b38	3.3e-5	-0.23	Colon - Transverse
ENSG00000161203.13	AP2M1	chr3_184257921_G_T_b38	5.1e-5	-0.15	Whole Blood
ENSG00000161204.11	ABCF3	chr3_184257921_G_T_b38	7.2e-5	0.21	Skin - Sun Exposed (Lower leg)

p-value, eQTL association p value; NES, normalized effect size.

Table S18. eQTL statistics for rs13132853 from GTEx v8 portal.

Gencode Id	Gene Symbol	Variant Id	P-Value	NES	Tissue
ENSG00000231160.9	KLF3-AS1	chr4_38678394_A_G_b38	9.1e-6	-0.17	Whole Blood
ENSG00000231160.9	KLF3-AS1	chr4_38678394_A_G_b38	1.4e-5	-0.15	Artery - Tibial
ENSG00000231160.9	KLF3-AS1	chr4_38678394_A_G_b38	3.2e-5	-0.36	Brain - Cerebellum
ENSG00000174125.7	TLR1	chr4_38678394_A_G_b38	3.5e-5	0.15	Esophagus - Mucosa
ENSG00000231160.9	KLF3-AS1	chr4_38678394_A_G_b38	1.4e-4	-0.16	Muscle - Skeletal

p-value, eQTL association p value; NES, normalized effect size.

Table S19. eQTL statistics for rs2790102 from GTEx v8 portal.

Gencode Id	Gene Symbol	Variant Id	P-Value	NES	Tissue
ENSG00000124813.20	RUNX2	chr6_45464477_G_A_b38	7.9e-12	-0.54	Brain - Putamen (basal ganglia)
ENSG00000124813.20	RUNX2	chr6_45464477_G_A_b38	6.2e-10	-0.39	Brain - Caudate (basal ganglia)
ENSG00000124813.20	RUNX2	chr6_45464477_G_A_b38	1.1e-6	-0.35	Brain - Hippocampus
ENSG00000124813.20	RUNX2	chr6_45464477_G_A_b38	7.4e-6	-0.29	Brain - Spinal cord (cervical c-1)

p-value, eQTL association p value; NES, normalized effect size.

Table S20. eQTL statistics for rs17203398 from GTEx v8 portal.

Gencode Id	Gene Symbol	Variant Id	P-Value	NES	Tissue
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	3.3e-32	0.40	Artery - Tibial
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	4.6e-16	0.29	Adipose - Subcutaneous
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	4.4e-14	0.30	Artery - Aorta
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	7.7e-12	0.25	Skin - Sun Exposed (Lower leg)
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	2.7e-10	0.22	Esophagus - Muscularis
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	4.1e-10	0.20	Cells - Cultured fibroblasts
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	7.1e-9	0.22	Adipose - Visceral (Omentum)
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	1.3e-8	0.18	Thyroid
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	1.4e-7	0.21	Breast - Mammary Tissue
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	3.0e-7	0.20	Skin - Not Sun Exposed (Suprapubic)
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	1.0e-6	0.16	Muscle - Skeletal
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	2.9e-6	0.25	Adrenal Gland
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	7.6e-6	-0.38	Liver
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	9.6e-6	0.19	Esophagus - Gastroesophageal Junction
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	2.1e-5	0.21	Heart - Left Ventricle
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	2.8e-5	0.17	Heart - Atrial Appendage
ENSG00000054983.16	GALC	chr14_87983503_G_C_b38	4.0e-5	0.18	Artery - Coronary

p-value, eQTL association p value; NES, normalized effect size.

Table S21. eQTL statistics for rs2106786 from GTEx v8 portal.

Gencode Id	Gene Symbol	P-Value	NES	Tissue
ENSG00000204650.14	LINC02210	2.5e-219	1.2	Skin - Sun Exposed (Lower leg)
ENSG00000204650.14	LINC02210	1.4e-207	1.3	Adipose - Subcutaneous
ENSG00000214425.7	LRRC37A4P	2.2e-193	-1.2	Whole Blood
ENSG00000204650.14	LINC02210	1.3e-190	1.2	Nerve - Tibial
ENSG00000204650.14	LINC02210	1.1e-182	1.3	Artery - Tibial
ENSG00000204650.14	LINC02210	2.3e-180	1.2	Skin - Not Sun Exposed (Suprapubic)
ENSG00000204650.14	LINC02210	6.4e-178	1.1	Adipose - Visceral (Omentum)
ENSG00000214425.7	LRRC37A4P	2.3e-172	-1.3	Thyroid
ENSG00000204650.14	LINC02210	3.9e-169	1.0	Thyroid
ENSG00000204650.14	LINC02210	7.9e-166	1.3	Cells - Cultured fibroblasts
ENSG00000204650.14	LINC02210	4.3e-164	1.2	Lung
ENSG00000214425.7	LRRC37A4P	5.8e-162	-1.2	Skin - Sun Exposed (Lower leg)
ENSG00000204650.14	LINC02210	1.7e-161	1.3	Esophagus - Muscularis
ENSG00000214401.4	KANSL1-AS1	6.2e-158	1.2	Muscle - Skeletal
ENSG00000262539.1	RP11-259G18.3	5.0e-154	1.3	Whole Blood
ENSG00000263503.1	MAPK8IP1P2	1.5e-153	1.2	Muscle - Skeletal
ENSG00000214425.7	LRRC37A4P	3.4e-151	-1.2	Lung
ENSG00000214401.4	KANSL1-AS1	1.3e-150	1.3	Whole Blood
ENSG00000214425.7	LRRC37A4P	7.3e-150	-1.2	Adipose - Subcutaneous
ENSG00000263503.1	MAPK8IP1P2	1.2e-145	1.2	Whole Blood
ENSG00000214425.7	LRRC37A4P	1.2e-144	-1.3	Nerve - Tibial
ENSG00000204650.14	LINC02210	1.8e-144	1.3	Artery - Aorta
ENSG00000214401.4	KANSL1-AS1	9.6e-144	1.2	Artery - Tibial
ENSG00000214425.7	LRRC37A4P	8.8e-141	-1.2	Skin - Not Sun Exposed (Suprapubic)
ENSG00000204650.14	LINC02210	1.9e-140	0.83	Muscle - Skeletal
ENSG00000262539.1	RP11-259G18.3	1.9e-138	1.3	Artery - Tibial
ENSG00000264070.1	DND1P1	1.3e-136	1.3	Muscle - Skeletal
ENSG00000214401.4	KANSL1-AS1	2.4e-134	1.3	Skin - Sun Exposed (Lower leg)
ENSG00000262539.1	RP11-259G18.3	1.5e-131	1.3	Skin - Sun Exposed (Lower leg)
ENSG00000262539.1	RP11-259G18.3	8.6e-130	1.2	Muscle - Skeletal
ENSG00000204650.14	LINC02210	3.8e-129	0.99	Breast - Mammary Tissue
ENSG00000262500.1	MAPK8IP1P1	1.9e-126	1.1	Whole Blood
ENSG00000262539.1	RP11-259G18.3	3.1e-126	1.3	Adipose - Subcutaneous
ENSG00000214401.4	KANSL1-AS1	1.3e-125	1.2	Thyroid
ENSG00000262539.1	RP11-259G18.3	4.6e-123	1.2	Cells - Cultured fibroblasts
ENSG00000214425.7	LRRC37A4P	2.7e-121	-1.3	Esophagus - Muscularis
ENSG00000214401.4	KANSL1-AS1	5.8e-121	1.2	Adipose - Subcutaneous
ENSG00000214401.4	KANSL1-AS1	8.6e-121	1.3	Skin - Not Sun Exposed (Suprapubic)
ENSG00000262500.1	MAPK8IP1P1	3.6e-119	1.1	Skin - Sun Exposed (Lower leg)
ENSG00000262539.1	RP11-259G18.3	3.7e-119	1.3	Thyroid

ENSG00000263503.1	MAPK8IP1P2	4.9e-119	1.2	Skin - Sun Exposed (Lower leg)
ENSG00000214425.7	LRRC37A4P	5.7e-119	-1.1	Artery - Tibial
ENSG00000238083.7	LRRC37A2	1.6e-118	1.2	Adipose - Subcutaneous
ENSG00000204650.14	LINC02210	2.7e-117	0.88	Colon - Transverse
ENSG00000186868.15	MAPT	3.7e-117	1.2	Testis
ENSG00000214425.7	LRRC37A4P	2.2e-116	-1.1	Cells - Cultured fibroblasts
ENSG00000264070.1	DND1P1	1.9e-115	1.3	Skin - Sun Exposed (Lower leg)
ENSG00000263503.1	MAPK8IP1P2	1.7e-114	1.2	Adipose - Subcutaneous
ENSG00000214425.7	LRRC37A4P	1.1e-113	-1.2	Adipose - Visceral (Omentum)
ENSG00000238083.7	LRRC37A2	1.2e-113	1.1	Thyroid
ENSG00000214425.7	LRRC37A4P	1.3e-113	-1.2	Colon - Transverse
ENSG00000214401.4	KANSL1-AS1	2.6e-113	1.2	Nerve - Tibial
ENSG00000264070.1	DND1P1	9.7e-113	1.3	Whole Blood
ENSG00000214401.4	KANSL1-AS1	1.8e-112	1.3	Adipose - Visceral (Omentum)
ENSG00000204650.14	LINC02210	1.9e-112	1.2	Colon - Sigmoid
ENSG00000214401.4	KANSL1-AS1	2.9e-112	1.2	Cells - Cultured fibroblasts
ENSG00000262539.1	RP11-259G18.3	1.2e-110	1.3	Skin - Not Sun Exposed (Suprapubic)
ENSG00000214401.4	KANSL1-AS1	1.9e-110	1.3	Lung
ENSG00000263503.1	MAPK8IP1P2	4.9e-110	1.2	Artery - Tibial
ENSG00000263503.1	MAPK8IP1P2	9.3e-110	1.2	Thyroid
ENSG00000263503.1	MAPK8IP1P2	2.3e-109	1.2	Cells - Cultured fibroblasts
ENSG00000263503.1	MAPK8IP1P2	4.9e-109	1.2	Skin - Not Sun Exposed (Suprapubic)
ENSG00000238083.7	LRRC37A2	5.9e-108	1.2	Artery - Tibial
ENSG00000204650.14	LINC02210	1.2e-107	0.58	Whole Blood
ENSG00000204650.14	LINC02210	8.9e-107	1.1	Heart - Atrial Appendage
ENSG00000262500.1	MAPK8IP1P1	2.2e-106	1.2	Skin - Not Sun Exposed (Suprapubic)
ENSG00000214401.4	KANSL1-AS1	4.4e-104	1.2	Esophagus - Muscularis
ENSG00000280022.1	RP11-707O23.1	9.1e-104	1.1	Muscle - Skeletal
ENSG00000262539.1	RP11-259G18.3	1.1e-103	1.2	Esophagus - Mucosa
ENSG00000262539.1	RP11-259G18.3	1.8e-103	1.3	Nerve - Tibial
ENSG00000204650.14	LINC02210	6.5e-102	1.2	Esophagus - Gastroesophageal Junction
ENSG00000262500.1	MAPK8IP1P1	9.7e-102	1.1	Thyroid
ENSG00000264070.1	DND1P1	1.8e-101	1.3	Artery - Tibial
ENSG00000264070.1	DND1P1	1.3e-100	1.3	Adipose - Subcutaneous
ENSG00000262539.1	RP11-259G18.3	5.6e-99	1.3	Adipose - Visceral (Omentum)
ENSG00000264070.1	DND1P1	4.3e-98	1.2	Thyroid
ENSG00000263503.1	MAPK8IP1P2	8.1e-97	1.2	Adipose - Visceral (Omentum)
ENSG00000214401.4	KANSL1-AS1	3.6e-96	1.2	Esophagus - Mucosa
ENSG00000263503.1	MAPK8IP1P2	4.0e-96	1.2	Nerve - Tibial
ENSG00000238083.7	LRRC37A2	7.3e-96	1.2	Nerve - Tibial
ENSG00000264070.1	DND1P1	8.7e-96	1.3	Skin - Not Sun Exposed (Suprapubic)
ENSG00000263503.1	MAPK8IP1P2	6.1e-95	1.2	Esophagus - Mucosa

ENSG00000214425.7	LRRC37A4P	5.1e-94	-1.1	Breast - Mammary Tissue
ENSG00000214425.7	LRRC37A4P	3.0e-93	-1.2	Artery - Aorta
ENSG00000204650.14	LINC02210	4.4e-93	0.59	Esophagus - Mucosa
ENSG00000262539.1	RP11-259G18.3	6.7e-93	1.2	Lung
ENSG00000262500.1	MAPK8IP1P1	1.7e-92	1.1	Lung
ENSG00000262539.1	RP11-259G18.3	1.4e-91	1.2	Esophagus - Muscularis
ENSG00000238083.7	LRRC37A2	6.8e-91	1.1	Esophagus - Muscularis
ENSG00000264070.1	DND1P1	3.5e-90	1.2	Nerve - Tibial
ENSG00000263503.1	MAPK8IP1P2	2.1e-89	1.2	Lung
ENSG00000204650.14	LINC02210	8.7e-89	1.1	Heart - Left Ventricle
ENSG00000214425.7	LRRC37A4P	2.5e-87	-1.3	Esophagus - Gastroesophageal Junction
ENSG00000214401.4	KANSL1-AS1	2.6e-86	1.2	Artery - Aorta
ENSG00000204650.14	LINC02210	3.3e-86	1.1	Stomach
ENSG00000214425.7	LRRC37A4P	6.7e-86	-1.0	Esophagus - Mucosa
ENSG00000280022.1	RP11-707O23.1	8.4e-86	1.2	Adipose - Subcutaneous
ENSG00000214425.7	LRRC37A4P	1.0e-85	-1.2	Colon - Sigmoid
ENSG00000238083.7	LRRC37A2	1.4e-85	0.99	Adipose - Visceral (Omentum)
ENSG00000264070.1	DND1P1	1.4e-85	1.4	Adipose - Visceral (Omentum)
ENSG00000263503.1	MAPK8IP1P2	2.0e-85	0.97	Testis
ENSG00000264070.1	DND1P1	2.1e-85	1.3	Lung
ENSG00000263503.1	MAPK8IP1P2	2.7e-85	1.2	Esophagus - Muscularis
ENSG00000280022.1	RP11-707O23.1	2.2e-84	1.2	Artery - Tibial
ENSG00000264070.1	DND1P1	5.2e-84	1.2	Cells - Cultured fibroblasts
ENSG00000214401.4	KANSL1-AS1	6.3e-84	1.2	Breast - Mammary Tissue
ENSG00000204650.14	LINC02210	6.4e-83	1.2	Pancreas
ENSG00000214425.7	LRRC37A4P	9.4e-83	-1.2	Heart - Atrial Appendage
ENSG00000264070.1	DND1P1	1.2e-82	1.3	Breast - Mammary Tissue
ENSG00000238083.7	LRRC37A2	2.3e-82	0.95	Lung
ENSG00000214401.4	KANSL1-AS1	8.1e-82	1.2	Esophagus - Gastroesophageal Junction
ENSG00000238083.7	LRRC37A2	1.4e-81	0.98	Skin - Sun Exposed (Lower leg)
ENSG00000262500.1	MAPK8IP1P1	3.2e-81	1.1	Adipose - Visceral (Omentum)
ENSG00000214401.4	KANSL1-AS1	1.3e-80	1.2	Heart - Left Ventricle
ENSG00000214401.4	KANSL1-AS1	5.1e-80	1.2	Colon - Transverse
ENSG00000262539.1	RP11-259G18.3	7.9e-80	1.3	Breast - Mammary Tissue
ENSG00000262539.1	RP11-259G18.3	1.6e-79	1.2	Artery - Aorta
ENSG00000264070.1	DND1P1	8.4e-79	1.2	Esophagus - Muscularis
ENSG00000262539.1	RP11-259G18.3	9.5e-79	1.2	Heart - Left Ventricle
ENSG00000263503.1	MAPK8IP1P2	3.4e-76	1.2	Colon - Transverse
ENSG00000214425.7	LRRC37A4P	6.3e-76	-1.3	Pituitary
ENSG00000262500.1	MAPK8IP1P1	1.3e-75	0.98	Esophagus - Mucosa
ENSG00000214401.4	KANSL1-AS1	1.7e-75	1.2	Heart - Atrial Appendage
ENSG00000214425.7	LRRC37A4P	1.9e-75	-0.95	Muscle - Skeletal

ENSG00000238083.7	LRRC37A2	2.8e-75	1.1	Artery - Aorta
ENSG00000204650.14	LINC02210	6.5e-74	0.72	Testis
ENSG00000262539.1	RP11-259G18.3	2.1e-73	1.3	Heart - Atrial Appendage
ENSG00000263503.1	MAPK8IP1P2	2.9e-73	1.2	Heart - Left Ventricle
ENSG00000264070.1	DND1P1	3.0e-73	1.3	Colon - Transverse
ENSG00000280022.1	RP11-707O23.1	4.3e-73	1.1	Skin - Sun Exposed (Lower leg)
ENSG00000238083.7	LRRC37A2	5.2e-73	0.98	Muscle - Skeletal
ENSG00000186868.15	MAPT	7.5e-73	0.69	Esophagus - Mucosa
ENSG00000262539.1	RP11-259G18.3	8.4e-73	1.4	Testis
ENSG00000263503.1	MAPK8IP1P2	9.0e-73	1.2	Breast - Mammary Tissue
ENSG00000214401.4	KANSL1-AS1	1.3e-72	1.2	Colon - Sigmoid
ENSG00000214425.7	LRRC37A4P	3.6e-72	-1.3	Pancreas
ENSG00000263503.1	MAPK8IP1P2	8.4e-71	1.1	Artery - Aorta
ENSG00000280022.1	RP11-707O23.1	1.1e-70	0.96	Whole Blood
ENSG00000204650.14	LINC02210	3.0e-70	1.1	Prostate
ENSG00000262539.1	RP11-259G18.3	1.7e-69	1.3	Colon - Sigmoid
ENSG00000204650.14	LINC02210	6.5e-69	1.3	Adrenal Gland
ENSG00000214425.7	LRRC37A4P	1.2e-68	-1.1	Heart - Left Ventricle
ENSG00000238083.7	LRRC37A2	1.4e-68	1.1	Breast - Mammary Tissue
ENSG00000262539.1	RP11-259G18.3	1.4e-68	1.2	Colon - Transverse
ENSG00000280022.1	RP11-707O23.1	1.6e-68	1.1	Skin - Not Sun Exposed (Suprapubic)
ENSG00000214425.7	LRRC37A4P	1.9e-68	-1.0	Stomach
ENSG00000204650.14	LINC02210	7.7e-68	1.1	Pituitary
ENSG00000204650.14	LINC02210	1.3e-67	1.3	Spleen
ENSG00000214425.7	LRRC37A4P	6.6e-67	-1.1	Testis
ENSG00000280022.1	RP11-707O23.1	6.8e-67	1.1	Esophagus - Muscularis
ENSG00000238083.7	LRRC37A2	7.2e-67	0.95	Esophagus - Mucosa
ENSG00000262539.1	RP11-259G18.3	1.1e-66	1.2	Esophagus - Gastroesophageal Junction
ENSG00000214401.4	KANSL1-AS1	1.2e-66	1.3	Stomach
ENSG00000263503.1	MAPK8IP1P2	1.7e-66	1.2	Heart - Atrial Appendage
ENSG00000262500.1	MAPK8IP1P1	2.6e-66	1.1	Colon - Transverse
ENSG00000214401.4	KANSL1-AS1	4.8e-66	1.3	Testis
ENSG00000262500.1	MAPK8IP1P1	9.1e-66	1.3	Testis
ENSG00000262539.1	RP11-259G18.3	1.0e-65	1.3	Stomach
ENSG00000264070.1	DND1P1	1.2e-65	1.3	Heart - Left Ventricle
ENSG00000262500.1	MAPK8IP1P1	2.9e-65	0.99	Esophagus - Muscularis
ENSG00000238083.7	LRRC37A2	4.7e-65	0.91	Skin - Not Sun Exposed (Suprapubic)
ENSG00000204650.14	LINC02210	8.9e-65	1.3	Artery - Coronary
ENSG00000262500.1	MAPK8IP1P1	2.5e-64	1.2	Colon - Sigmoid
ENSG00000263503.1	MAPK8IP1P2	2.9e-63	1.2	Stomach
ENSG00000261575.2	RP11-259G18.1	7.1e-63	1.2	Testis
ENSG00000263503.1	MAPK8IP1P2	9.6e-63	1.2	Esophagus - Gastroesophageal Junction

ENSG00000264070.1	DND1P1	3.5e-62	1.3	Colon - Sigmoid
ENSG00000280022.1	RP11-707O23.1	4.4e-61	1.0	Nerve - Tibial
ENSG00000238083.7	LRRC37A2	1.2e-60	1.0	Esophagus - Gastroesophageal Junction
ENSG00000238083.7	LRRC37A2	1.9e-60	0.84	Whole Blood
ENSG00000263503.1	MAPK8IP1P2	1.3e-59	1.2	Colon - Sigmoid
ENSG00000262500.1	MAPK8IP1P1	1.2e-58	1.1	Breast - Mammary Tissue
ENSG00000264070.1	DND1P1	2.6e-58	1.2	Artery - Aorta
ENSG00000262500.1	MAPK8IP1P1	4.3e-57	1.1	Esophagus - Gastroesophageal Junction
ENSG00000214425.7	LRRC37A4P	7.4e-57	-1.2	Spleen
ENSG00000238083.7	LRRC37A2	1.1e-56	1.0	Heart - Left Ventricle
ENSG00000214425.7	LRRC37A4P	6.7e-56	-1.0	Brain - Nucleus accumbens (basal ganglia)
ENSG00000280022.1	RP11-707O23.1	1.7e-55	1.1	Adipose - Visceral (Omentum)
ENSG00000264070.1	DND1P1	3.8e-55	0.99	Esophagus - Mucosa
ENSG00000280022.1	RP11-707O23.1	6.7e-55	1.1	Esophagus - Gastroesophageal Junction
ENSG00000238083.7	LRRC37A2	8.6e-54	1.1	Colon - Sigmoid
ENSG00000238083.7	LRRC37A2	1.8e-53	1.1	Colon - Transverse
ENSG00000238083.7	LRRC37A2	1.9e-53	0.98	Heart - Atrial Appendage
ENSG00000262539.1	RP11-259G18.3	1.9e-53	1.2	Liver
ENSG00000204650.14	LINC02210	2.5e-52	1.2	Ovary
ENSG00000204650.14	LINC02210	3.3e-52	0.87	Small Intestine - Terminal Ileum
ENSG00000262500.1	MAPK8IP1P1	3.5e-52	0.91	Nerve - Tibial
ENSG00000262539.1	RP11-259G18.3	4.0e-52	1.3	Pancreas
ENSG00000238083.7	LRRC37A2	1.2e-51	1.0	Stomach
ENSG00000238083.7	LRRC37A2	1.4e-51	1.2	Brain - Cerebellum
ENSG00000214425.7	LRRC37A4P	1.9e-51	-1.0	Brain - Caudate (basal ganglia)
ENSG00000264070.1	DND1P1	2.0e-51	1.2	Stomach
ENSG00000214425.7	LRRC37A4P	2.1e-51	-1.2	Brain - Cortex
ENSG00000214425.7	LRRC37A4P	1.7e-50	-1.2	Artery - Coronary
ENSG00000264070.1	DND1P1	2.2e-50	1.2	Esophagus - Gastroesophageal Junction
ENSG00000214425.7	LRRC37A4P	2.2e-50	-1.0	Prostate
ENSG00000214401.4	KANSL1-AS1	2.5e-50	1.2	Pituitary
ENSG00000264070.1	DND1P1	2.9e-50	1.2	Heart - Atrial Appendage
ENSG00000238083.7	LRRC37A2	4.5e-50	1.2	Pituitary
ENSG00000264070.1	DND1P1	6.3e-50	1.2	Prostate
ENSG00000214401.4	KANSL1-AS1	7.0e-50	1.2	Liver
ENSG00000261575.2	RP11-259G18.1	7.1e-50	1.2	Brain - Cerebellum
ENSG00000214401.4	KANSL1-AS1	1.3e-49	1.1	Pancreas
ENSG00000280022.1	RP11-707O23.1	3.2e-49	1.2	Colon - Sigmoid
ENSG00000214401.4	KANSL1-AS1	4.6e-49	1.2	Brain - Cortex
ENSG00000280022.1	RP11-707O23.1	1.4e-47	1.0	Lung
ENSG00000185829.17	ARL17A	2.1e-47	0.88	Thyroid
ENSG00000262539.1	RP11-259G18.3	2.3e-47	1.3	Pituitary

ENSG00000280022.1	RP11-707O23.1	2.7e-47	1.1	Artery - Aorta
ENSG00000238083.7	LRRC37A2	3.6e-47	0.93	Cells - Cultured fibroblasts
ENSG00000263503.1	MAPK8IP1P2	4.4e-47	1.2	Liver
ENSG00000214401.4	KANSL1-AS1	1.4e-46	1.3	Adrenal Gland
ENSG00000280022.1	RP11-707O23.1	1.6e-46	1.1	Breast - Mammary Tissue
ENSG00000214425.7	LRRC37A4P	1.6e-46	-1.2	Brain - Frontal Cortex (BA9)
ENSG00000214425.7	LRRC37A4P	3.9e-46	-1.2	Cells - EBV-transformed lymphocytes
ENSG00000262539.1	RP11-259G18.3	4.2e-46	1.3	Prostate
ENSG00000262539.1	RP11-259G18.3	5.0e-46	1.4	Spleen
ENSG00000214425.7	LRRC37A4P	1.2e-45	-1.2	Brain - Hypothalamus
ENSG00000262539.1	RP11-259G18.3	1.3e-45	1.3	Adrenal Gland
ENSG00000214425.7	LRRC37A4P	1.5e-45	-1.1	Brain - Cerebellum
ENSG00000214425.7	LRRC37A4P	2.7e-45	-0.92	Small Intestine - Terminal Ileum
ENSG00000262500.1	MAPK8IP1P1	4.8e-45	1.2	Spleen
ENSG00000261575.2	RP11-259G18.1	5.5e-45	0.78	Skin - Sun Exposed (Lower leg)
ENSG00000262539.1	RP11-259G18.3	8.0e-45	1.2	Artery - Coronary
ENSG00000204652.6	RPS26P8	1.0e-44	0.98	Nerve - Tibial
ENSG00000238083.7	LRRC37A2	2.0e-44	1.2	Brain - Nucleus accumbens (basal ganglia)
ENSG00000214425.7	LRRC37A4P	3.2e-44	-1.2	Brain - Cerebellar Hemisphere
ENSG00000262539.1	RP11-259G18.3	4.6e-44	1.2	Brain - Cerebellum
ENSG00000214401.4	KANSL1-AS1	5.3e-44	1.2	Prostate
ENSG00000280022.1	RP11-707O23.1	6.2e-44	1.1	Colon - Transverse
ENSG00000176681.14	LRRC37A	9.4e-44	0.95	Nerve - Tibial
ENSG00000214401.4	KANSL1-AS1	1.1e-43	1.1	Brain - Nucleus accumbens (basal ganglia)
ENSG00000176681.14	LRRC37A	1.1e-43	0.87	Artery - Tibial
ENSG00000214425.7	LRRC37A4P	1.4e-43	-1.2	Brain - Hippocampus
ENSG00000214401.4	KANSL1-AS1	1.6e-43	1.2	Artery - Coronary
ENSG00000263503.1	MAPK8IP1P2	4.0e-43	1.2	Pituitary
ENSG00000214425.7	LRRC37A4P	4.4e-43	-1.2	Adrenal Gland
ENSG00000280022.1	RP11-707O23.1	5.6e-43	0.82	Cells - Cultured fibroblasts
ENSG00000238083.7	LRRC37A2	2.1e-42	1.3	Brain - Cerebellar Hemisphere
ENSG00000214401.4	KANSL1-AS1	2.3e-42	1.2	Brain - Cerebellum
ENSG00000238083.7	LRRC37A2	6.8e-42	1.2	Brain - Cortex
ENSG00000264070.1	DND1P1	8.8e-42	1.2	Pituitary
ENSG00000108379.9	WNT3	1.7e-41	0.66	Thyroid
ENSG00000264070.1	DND1P1	2.6e-41	1.2	Small Intestine - Terminal Ileum
ENSG00000263503.1	MAPK8IP1P2	3.0e-41	1.2	Prostate
ENSG00000214401.4	KANSL1-AS1	3.4e-41	1.2	Brain - Putamen (basal ganglia)
ENSG00000238083.7	LRRC37A2	6.6e-41	1.3	Brain - Hypothalamus
ENSG00000238083.7	LRRC37A2	7.5e-41	1.2	Brain - Caudate (basal ganglia)
ENSG00000261575.2	RP11-259G18.1	9.8e-41	1.2	Brain - Cerebellar Hemisphere
ENSG00000264070.1	DND1P1	1.1e-40	1.3	Liver

ENSG00000214401.4	KANSL1-AS1	1.1e-40	1.3	Small Intestine - Terminal Ileum
ENSG00000264070.1	DND1P1	1.4e-40	1.3	Adrenal Gland
ENSG00000280022.1	RP11-707O23.1	1.8e-40	1.1	Heart - Atrial Appendage
ENSG00000263503.1	MAPK8IP1P2	2.6e-40	1.3	Adrenal Gland
ENSG00000225190.10	PLEKHM1	5.5e-40	-1.0	Brain - Cerebellum
ENSG00000263503.1	MAPK8IP1P2	8.4e-40	1.2	Spleen
ENSG00000262539.1	RP11-259G18.3	1.0e-39	1.2	Brain - Nucleus accumbens (basal ganglia)
ENSG00000262539.1	RP11-259G18.3	1.5e-39	1.3	Brain - Caudate (basal ganglia)
ENSG00000262539.1	RP11-259G18.3	1.6e-39	1.2	Brain - Cortex
ENSG00000204650.14	LINC02210	2.3e-39	1.1	Uterus
ENSG00000262500.1	MAPK8IP1P1	2.6e-39	1.2	Prostate
ENSG00000262500.1	MAPK8IP1P1	2.7e-39	1.2	Pituitary
ENSG00000280022.1	RP11-707O23.1	6.6e-39	1.0	Heart - Left Ventricle
ENSG00000176681.14	LRRC37A	6.8e-39	0.76	Skin - Sun Exposed (Lower leg)
ENSG00000262539.1	RP11-259G18.3	9.8e-39	1.3	Small Intestine - Terminal Ileum
ENSG00000263503.1	MAPK8IP1P2	1.1e-38	1.3	Small Intestine - Terminal Ileum
ENSG00000214425.7	LRRC37A4P	1.8e-38	-1.2	Brain - Anterior cingulate cortex (BA24)
ENSG00000280022.1	RP11-707O23.1	3.7e-38	1.1	Pancreas
ENSG00000280022.1	RP11-707O23.1	4.6e-38	0.80	Thyroid
ENSG00000214401.4	KANSL1-AS1	7.2e-38	1.4	Minor Salivary Gland
ENSG00000214425.7	LRRC37A4P	1.2e-37	-1.1	Liver
ENSG00000263503.1	MAPK8IP1P2	1.3e-37	1.2	Artery - Coronary
ENSG00000214401.4	KANSL1-AS1	1.4e-37	1.2	Spleen
ENSG00000280022.1	RP11-707O23.1	2.4e-37	1.0	Stomach
ENSG00000280022.1	RP11-707O23.1	2.5e-37	1.2	Liver
ENSG00000204650.14	LINC02210	4.8e-37	0.69	Brain - Nucleus accumbens (basal ganglia)
ENSG00000280022.1	RP11-707O23.1	5.1e-37	1.0	Testis
ENSG00000204650.14	LINC02210	5.3e-37	0.89	Brain - Caudate (basal ganglia)
ENSG00000262500.1	MAPK8IP1P1	6.2e-37	1.2	Brain - Cerebellum
ENSG00000238083.7	LRRC37A2	6.9e-37	1.1	Prostate
ENSG00000214425.7	LRRC37A4P	9.4e-37	-1.4	Ovary
ENSG00000214401.4	KANSL1-AS1	1.1e-36	1.1	Brain - Caudate (basal ganglia)
ENSG00000238083.7	LRRC37A2	1.6e-36	1.0	Artery - Coronary
ENSG00000185829.17	ARL17A	1.8e-36	1.1	Brain - Cerebellum
ENSG00000263503.1	MAPK8IP1P2	1.9e-36	1.2	Brain - Cerebellum
ENSG00000264070.1	DND1P1	1.9e-36	1.3	Spleen
ENSG00000264070.1	DND1P1	8.6e-36	1.1	Pancreas
ENSG00000238083.7	LRRC37A2	1.2e-35	1.0	Pancreas
ENSG00000263503.1	MAPK8IP1P2	1.3e-35	1.0	Pancreas
ENSG00000238083.7	LRRC37A2	1.5e-35	1.2	Brain - Putamen (basal ganglia)
ENSG00000262539.1	RP11-259G18.3	3.4e-35	1.2	Brain - Putamen (basal ganglia)
ENSG00000159314.11	ARHGAP27	3.9e-35	0.39	Skin - Not Sun Exposed (Suprapubic)

ENSG00000238083.7	LRRC37A2	4.3e-35	1.2	Brain - Frontal Cortex (BA9)
ENSG00000238083.7	LRRC37A2	5.5e-35	1.1	Adrenal Gland
ENSG00000264070.1	DND1P1	6.2e-35	1.3	Artery - Coronary
ENSG00000261575.2	RP11-259G18.1	6.9e-35	0.73	Skin - Not Sun Exposed (Suprapubic)
ENSG00000263503.1	MAPK8IP1P2	1.5e-34	1.2	Brain - Cortex
ENSG00000204650.14	LINC02210	1.6e-34	1.3	Brain - Spinal cord (cervical c-1)
ENSG00000176681.14	LRRC37A	2.6e-34	0.80	Adipose - Subcutaneous
ENSG00000176681.14	LRRC37A	3.3e-34	0.74	Skin - Not Sun Exposed (Suprapubic)
ENSG00000228696.8	ARL17B	5.5e-34	1.1	Testis
ENSG00000262539.1	RP11-259G18.3	5.7e-34	1.3	Brain - Cerebellar Hemisphere
ENSG00000262500.1	MAPK8IP1P1	7.2e-34	1.1	Liver
ENSG00000262539.1	RP11-259G18.3	7.9e-34	1.3	Cells - EBV-transformed lymphocytes
ENSG00000262500.1	MAPK8IP1P1	8.8e-34	1.1	Small Intestine - Terminal Ileum
ENSG00000264070.1	DND1P1	9.1e-34	1.2	Brain - Cortex
ENSG00000238083.7	LRRC37A2	9.3e-34	1.2	Brain - Hippocampus
ENSG00000214401.4	KANSL1-AS1	2.1e-33	1.1	Brain - Cerebellar Hemisphere
ENSG00000214401.4	KANSL1-AS1	2.8e-33	1.2	Vagina
ENSG00000225190.10	PLEKHM1	3.9e-33	0.31	Muscle - Skeletal
ENSG00000159314.11	ARHGAP27	4.3e-33	0.40	Skin - Sun Exposed (Lower leg)
ENSG00000185829.17	ARL17A	8.2e-33	1.1	Brain - Cerebellar Hemisphere
ENSG00000214401.4	KANSL1-AS1	8.3e-33	1.1	Brain - Frontal Cortex (BA9)
ENSG00000186868.15	MAPT	3.4e-32	0.59	Lung
ENSG00000214425.7	LRRC37A4P	3.4e-32	-1.3	Brain - Spinal cord (cervical c-1)
ENSG00000262539.1	RP11-259G18.3	4.1e-32	1.4	Minor Salivary Gland
ENSG00000204650.14	LINC02210	6.3e-32	0.81	Brain - Hypothalamus
ENSG00000263503.1	MAPK8IP1P2	1.4e-31	1.2	Brain - Nucleus accumbens (basal ganglia)
ENSG00000262500.1	MAPK8IP1P1	1.6e-31	1.1	Brain - Cortex
ENSG00000214425.7	LRRC37A4P	1.9e-31	-1.1	Brain - Putamen (basal ganglia)
ENSG00000280022.1	RP11-707O23.1	7.8e-31	0.72	Esophagus - Mucosa
ENSG00000262881.1	RP11-669E14.4	1.6e-30	-0.62	Liver
ENSG00000280022.1	RP11-707O23.1	1.8e-30	1.1	Adrenal Gland
ENSG00000262500.1	MAPK8IP1P1	2.2e-30	1.2	Brain - Cerebellar Hemisphere
ENSG00000204650.14	LINC02210	7.1e-30	0.79	Brain - Hippocampus
ENSG00000214425.7	LRRC37A4P	1.1e-29	-1.1	Brain - Amygdala
ENSG00000238083.7	LRRC37A2	1.2e-29	1.2	Brain - Anterior cingulate cortex (BA24)
ENSG00000262500.1	MAPK8IP1P1	1.5e-29	1.1	Brain - Caudate (basal ganglia)
ENSG00000204650.14	LINC02210	2.0e-29	0.85	Brain - Putamen (basal ganglia)
ENSG00000238083.7	LRRC37A2	2.2e-29	1.1	Spleen
ENSG00000176681.14	LRRC37A	2.5e-29	0.75	Esophagus - Muscularis
ENSG00000262539.1	RP11-259G18.3	2.8e-29	1.1	Brain - Hypothalamus
ENSG00000232300.1	FAM215B	4.5e-29	0.72	Thyroid
ENSG00000204650.14	LINC02210	4.5e-29	0.69	Vagina

ENSG00000262539.1	RP11-259G18.3	5.4e-29	1.3	Uterus
ENSG00000204650.14	LINC02210	1.0e-28	0.75	Brain - Anterior cingulate cortex (BA24)
ENSG00000204650.14	LINC02210	1.0e-28	0.91	Brain - Amygdala
ENSG00000262539.1	RP11-259G18.3	1.0e-28	1.2	Brain - Frontal Cortex (BA9)
ENSG00000204650.14	LINC02210	1.3e-28	1.1	Minor Salivary Gland
ENSG00000214425.7	LRRC37A4P	1.4e-28	-1.3	Uterus
ENSG00000214401.4	KANSL1-AS1	1.4e-28	1.1	Brain - Hippocampus
ENSG00000264070.1	DND1P1	2.1e-28	1.2	Brain - Putamen (basal ganglia)
ENSG00000263503.1	MAPK8IP1P2	3.2e-28	1.1	Brain - Putamen (basal ganglia)
ENSG00000204650.14	LINC02210	4.6e-28	0.60	Brain - Cortex
ENSG00000263503.1	MAPK8IP1P2	5.3e-28	1.1	Cells - EBV-transformed lymphocytes
ENSG00000185829.17	ARL17A	5.5e-28	0.61	Muscle - Skeletal
ENSG00000280022.1	RP11-707O23.1	5.8e-28	1.1	Artery - Coronary
ENSG00000214425.7	LRRC37A4P	6.2e-28	-1.3	Brain - Substantia nigra
ENSG00000263503.1	MAPK8IP1P2	8.2e-28	1.2	Brain - Hypothalamus
ENSG00000262500.1	MAPK8IP1P1	8.3e-28	1.1	Brain - Nucleus accumbens (basal ganglia)
ENSG00000185829.17	ARL17A	8.9e-28	1.1	Brain - Cortex
ENSG00000263503.1	MAPK8IP1P2	9.2e-28	1.3	Minor Salivary Gland
ENSG00000263503.1	MAPK8IP1P2	1.1e-27	1.2	Brain - Cerebellar Hemisphere
ENSG00000214401.4	KANSL1-AS1	1.2e-27	1.1	Brain - Hypothalamus
ENSG00000263503.1	MAPK8IP1P2	1.2e-27	1.1	Brain - Caudate (basal ganglia)
ENSG00000225190.10	PLEKHM1	1.3e-27	-0.84	Brain - Cerebellar Hemisphere
ENSG00000262539.1	RP11-259G18.3	2.0e-27	1.2	Brain - Anterior cingulate cortex (BA24)
ENSG00000238083.7	LRRC37A2	2.4e-27	1.1	Small Intestine - Terminal Ileum
ENSG00000214401.4	KANSL1-AS1	2.9e-27	1.2	Uterus
ENSG00000261575.2	RP11-259G18.1	3.2e-27	0.72	Esophagus - Mucosa
ENSG00000214401.4	KANSL1-AS1	3.8e-27	1.2	Ovary
ENSG00000214401.4	KANSL1-AS1	4.5e-27	1.1	Cells - EBV-transformed lymphocytes
ENSG00000176681.14	LRRC37A	6.3e-27	1.0	Brain - Cerebellum
ENSG00000263503.1	MAPK8IP1P2	9.5e-27	1.2	Brain - Frontal Cortex (BA9)
ENSG00000262500.1	MAPK8IP1P1	1.7e-26	1.2	Brain - Hippocampus
ENSG00000262539.1	RP11-259G18.3	1.9e-26	1.3	Ovary
ENSG00000185829.17	ARL17A	3.2e-26	1.1	Brain - Frontal Cortex (BA9)
ENSG00000262539.1	RP11-259G18.3	3.2e-26	1.2	Brain - Hippocampus
ENSG00000204650.14	LINC02210	4.8e-26	0.56	Brain - Frontal Cortex (BA9)
ENSG00000262539.1	RP11-259G18.3	6.0e-26	1.1	Vagina
ENSG00000263503.1	MAPK8IP1P2	8.3e-26	1.2	Brain - Hippocampus
ENSG00000238083.7	LRRC37A2	9.0e-26	1.1	Brain - Amygdala
ENSG00000214401.4	KANSL1-AS1	1.0e-25	1.1	Brain - Anterior cingulate cortex (BA24)
ENSG00000185829.17	ARL17A	1.1e-25	0.99	Pituitary
ENSG00000261575.2	RP11-259G18.1	2.1e-25	0.68	Pancreas
ENSG00000263503.1	MAPK8IP1P2	2.2e-25	1.2	Brain - Anterior cingulate cortex (BA24)

ENSG00000176681.14	LRRC37A	2.8e-25	1.1	Brain - Cerebellar Hemisphere
ENSG00000186868.15	MAPT	5.4e-25	-0.67	Colon - Sigmoid
ENSG00000262500.1	MAPK8IP1P1	6.9e-25	1.1	Brain - Putamen (basal ganglia)
ENSG00000262500.1	MAPK8IP1P1	8.1e-25	1.0	Vagina
ENSG00000214425.7	LRRC37A4P	8.1e-25	-1.0	Vagina
ENSG00000263503.1	MAPK8IP1P2	9.3e-25	1.2	Brain - Amygdala
ENSG00000264070.1	DND1P1	1.8e-24	1.0	Brain - Caudate (basal ganglia)
ENSG00000225190.10	PLEKHM1	2.1e-24	-0.31	Esophagus - Mucosa
ENSG00000262500.1	MAPK8IP1P1	2.4e-24	0.97	Adrenal Gland
ENSG00000204650.14	LINC02210	2.9e-24	1.1	Brain - Substantia nigra
ENSG00000262500.1	MAPK8IP1P1	3.6e-24	1.2	Brain - Frontal Cortex (BA9)
ENSG00000185294.6	SPPL2C	5.5e-24	0.76	Brain - Cerebellum
ENSG00000238083.7	LRRC37A2	5.7e-24	0.89	Liver
ENSG00000264070.1	DND1P1	8.6e-24	1.1	Vagina
ENSG00000262500.1	MAPK8IP1P1	1.8e-23	1.2	Brain - Hypothalamus
ENSG00000185829.17	ARL17A	2.6e-23	0.99	Brain - Nucleus accumbens (basal ganglia)
ENSG00000262539.1	RP11-259G18.3	2.8e-23	1.2	Brain - Spinal cord (cervical c-1)
ENSG00000262500.1	MAPK8IP1P1	3.7e-23	1.2	Brain - Anterior cingulate cortex (BA24)
ENSG00000264070.1	DND1P1	5.2e-23	1.2	Ovary
ENSG00000120088.14	CRHR1	5.3e-23	0.44	Breast - Mammary Tissue
ENSG00000262539.1	RP11-259G18.3	6.8e-23	1.1	Brain - Amygdala
ENSG00000264070.1	DND1P1	7.4e-23	1.0	Brain - Nucleus accumbens (basal ganglia)
ENSG00000266918.1	RP11-798G7.8	8.9e-23	0.56	Muscle - Skeletal
ENSG00000262500.1	MAPK8IP1P1	1.3e-22	1.1	Ovary
ENSG00000184922.13	FMNL1	1.6e-22	-0.70	Brain - Cerebellum
ENSG00000261575.2	RP11-259G18.1	1.9e-22	0.70	Esophagus - Muscularis
ENSG00000263503.1	MAPK8IP1P2	2.8e-22	1.1	Vagina
ENSG00000232300.1	FAM215B	3.8e-22	0.85	Brain - Cerebellum
ENSG00000120071.13	KANSL1	3.9e-22	0.40	Muscle - Skeletal
ENSG00000214401.4	KANSL1-AS1	4.9e-22	1.1	Brain - Spinal cord (cervical c-1)
ENSG00000261575.2	RP11-259G18.1	1.1e-21	0.67	Nerve - Tibial
ENSG00000186868.15	MAPT	1.1e-21	0.49	Skin - Sun Exposed (Lower leg)
ENSG00000238083.7	LRRC37A2	1.1e-21	1.1	Brain - Spinal cord (cervical c-1)
ENSG00000214425.7	LRRC37A4P	1.1e-21	-1.0	Minor Salivary Gland
ENSG00000261575.2	RP11-259G18.1	1.4e-21	0.85	Brain - Nucleus accumbens (basal ganglia)
ENSG00000263503.1	MAPK8IP1P2	2.0e-21	1.1	Uterus
ENSG00000280022.1	RP11-707O23.1	2.1e-21	1.2	Ovary
ENSG00000176681.14	LRRC37A	2.6e-21	0.60	Adipose - Visceral (Omentum)
ENSG00000238083.7	LRRC37A2	2.8e-21	0.92	Ovary
ENSG00000214401.4	KANSL1-AS1	3.9e-21	1.0	Brain - Amygdala
ENSG00000264070.1	DND1P1	4.6e-21	1.0	Brain - Frontal Cortex (BA9)
ENSG00000185829.17	ARL17A	7.9e-21	1.0	Brain - Hypothalamus

ENSG00000264070.1	DND1P1	9.3e-21	1.0	Brain - Hippocampus
ENSG00000264070.1	DND1P1	9.3e-21	1.0	Brain - Hypothalamus
ENSG00000214425.7	LRRC37A4P	1.3e-20	-1.2	Kidney - Cortex
ENSG00000238083.7	LRRC37A2	1.4e-20	0.91	Uterus
ENSG00000264070.1	DND1P1	1.5e-20	1.1	Cells - EBV-transformed lymphocytes
ENSG00000120088.14	CRHR1	1.6e-20	0.51	Adipose - Visceral (Omentum)
ENSG00000238083.7	LRRC37A2	2.5e-20	1.2	Brain - Substantia nigra
ENSG00000264589.2	MAPT-AS1	2.9e-20	-0.70	Brain - Nucleus accumbens (basal ganglia)
ENSG00000185829.17	ARL17A	3.3e-20	0.69	Cells - Cultured fibroblasts
ENSG00000263503.1	MAPK8IP1P2	5.3e-20	1.1	Ovary
ENSG00000262500.1	MAPK8IP1P1	6.6e-20	1.1	Brain - Amygdala
ENSG00000185294.6	SPPL2C	8.3e-20	0.68	Brain - Cerebellar Hemisphere
ENSG00000264070.1	DND1P1	1.0e-19	-0.51	Testis
ENSG00000261575.2	RP11-259G18.1	1.4e-19	0.58	Thyroid
ENSG00000280022.1	RP11-707O23.1	1.5e-19	1.1	Uterus
ENSG00000261575.2	RP11-259G18.1	2.3e-19	0.77	Esophagus - Gastroesophageal Junction
ENSG00000238083.7	LRRC37A2	2.6e-19	0.89	Vagina
ENSG00000261575.2	RP11-259G18.1	3.6e-19	0.96	Brain - Cortex
ENSG00000238083.7	LRRC37A2	3.9e-19	1.1	Minor Salivary Gland
ENSG00000261575.2	RP11-259G18.1	4.5e-19	0.93	Brain - Caudate (basal ganglia)
ENSG00000184922.13	FMNL1	5.1e-19	-0.64	Brain - Cerebellar Hemisphere
ENSG00000176681.14	LRRC37A	5.8e-19	0.56	Lung
ENSG00000264070.1	DND1P1	6.1e-19	1.1	Brain - Anterior cingulate cortex (BA24)
ENSG00000280022.1	RP11-707O23.1	6.4e-19	0.78	Small Intestine - Terminal Ileum
ENSG00000262500.1	MAPK8IP1P1	1.8e-18	0.99	Uterus
ENSG00000176681.14	LRRC37A	2.9e-18	0.64	Artery - Aorta
ENSG00000159314.11	ARHGAP27	4.1e-18	0.43	Brain - Nucleus accumbens (basal ganglia)
ENSG00000176681.14	LRRC37A	4.2e-18	0.68	Breast - Mammary Tissue
ENSG00000120071.13	KANSL1	5.0e-18	0.35	Esophagus - Mucosa
ENSG00000204650.14	LINC02210	5.8e-18	1.1	Kidney - Cortex
ENSG00000204650.14	LINC02210	9.7e-18	0.72	Cells - EBV-transformed lymphocytes
ENSG00000108379.9	WNT3	1.1e-17	0.72	Pancreas
ENSG00000176681.14	LRRC37A	1.6e-17	0.65	Esophagus - Gastroesophageal Junction
ENSG00000214401.4	KANSL1-AS1	1.7e-17	1.0	Brain - Substantia nigra
ENSG00000274883.1	Metazoa_SRP	1.9e-17	-0.57	Liver
ENSG00000185829.17	ARL17A	2.0e-17	0.59	Esophagus - Muscularis
ENSG00000185829.17	ARL17A	2.2e-17	0.86	Brain - Caudate (basal ganglia)
ENSG00000120088.14	CRHR1	2.2e-17	0.50	Adipose - Subcutaneous
ENSG00000264070.1	DND1P1	3.0e-17	1.1	Brain - Spinal cord (cervical c-1)
ENSG00000185829.17	ARL17A	3.1e-17	0.88	Adrenal Gland
ENSG00000159314.11	ARHGAP27	3.6e-17	-0.17	Esophagus - Mucosa
ENSG00000264070.1	DND1P1	4.0e-17	0.71	Brain - Cerebellum

ENSG00000185829.17	ARL17A	1.1e-16	0.58	Adipose - Visceral (Omentum)
ENSG00000120071.13	KANSL1	1.6e-16	0.27	Cells - Cultured fibroblasts
ENSG00000280022.1	RP11-707O23.1	1.8e-16	0.76	Pituitary
ENSG00000263503.1	MAPK8IP1P2	2.0e-16	1.1	Brain - Spinal cord (cervical c-1)
ENSG00000264070.1	DND1P1	2.3e-16	0.97	Brain - Amygdala
ENSG00000186868.15	MAPT	2.3e-16	-0.48	Heart - Left Ventricle
ENSG00000214401.4	KANSL1-AS1	2.6e-16	1.1	Kidney - Cortex
ENSG00000185829.17	ARL17A	2.6e-16	0.53	Adipose - Subcutaneous
ENSG00000108379.9	WNT3	2.9e-16	0.54	Pituitary
ENSG00000266918.1	RP11-798G7.8	3.0e-16	0.55	Thyroid
ENSG00000261575.2	RP11-259G18.1	4.1e-16	0.74	Pituitary
ENSG00000120071.13	KANSL1	4.1e-16	0.31	Skin - Sun Exposed (Lower leg)
ENSG00000108379.9	WNT3	4.9e-16	0.34	Lung
ENSG00000176681.14	LRRC37A	5.0e-16	0.54	Thyroid
ENSG00000264070.1	DND1P1	5.4e-16	1.0	Uterus
ENSG00000176681.14	LRRC37A	8.1e-16	0.78	Artery - Coronary
ENSG00000185829.17	ARL17A	8.6e-16	0.69	Colon - Transverse
ENSG00000262500.1	MAPK8IP1P1	9.4e-16	1.2	Brain - Substantia nigra
ENSG00000261575.2	RP11-259G18.1	1.0e-15	0.68	Artery - Aorta
ENSG00000261575.2	RP11-259G18.1	1.4e-15	0.95	Brain - Putamen (basal ganglia)
ENSG00000261575.2	RP11-259G18.1	2.0e-15	0.89	Spleen
ENSG00000280022.1	RP11-707O23.1	2.1e-15	0.65	Prostate
ENSG00000176681.14	LRRC37A	2.6e-15	0.67	Colon - Sigmoid
ENSG00000280022.1	RP11-707O23.1	3.4e-15	0.92	Vagina
ENSG00000261575.2	RP11-259G18.1	4.7e-15	0.48	Adipose - Subcutaneous
ENSG00000185829.17	ARL17A	5.2e-15	0.55	Lung
ENSG00000108379.9	WNT3	6.2e-15	0.44	Esophagus - Muscularis
ENSG00000262500.1	MAPK8IP1P1	6.3e-15	1.1	Brain - Spinal cord (cervical c-1)
ENSG00000185829.17	ARL17A	7.5e-15	0.56	Esophagus - Mucosa
ENSG00000176681.14	LRRC37A	8.3e-15	0.47	Whole Blood
ENSG00000185829.17	ARL17A	9.2e-15	0.74	Pancreas
ENSG00000185829.17	ARL17A	9.4e-15	0.67	Stomach
ENSG00000261575.2	RP11-259G18.1	9.6e-15	0.68	Colon - Sigmoid
ENSG00000232300.1	FAM215B	1.2e-14	0.84	Brain - Cerebellar Hemisphere
ENSG00000261575.2	RP11-259G18.1	1.4e-14	0.60	Cells - Cultured fibroblasts
ENSG00000264589.2	MAPT-AS1	1.7e-14	-0.55	Brain - Hypothalamus
ENSG00000108379.9	WNT3	2.1e-14	0.50	Artery - Aorta
ENSG00000120071.13	KANSL1	2.1e-14	0.32	Skin - Not Sun Exposed (Suprapubic)
ENSG00000266918.1	RP11-798G7.8	2.4e-14	0.84	Brain - Cerebellar Hemisphere
ENSG00000176681.14	LRRC37A	3.0e-14	0.48	Muscle - Skeletal
ENSG00000264070.1	DND1P1	3.5e-14	1.1	Brain - Substantia nigra
ENSG00000261575.2	RP11-259G18.1	4.2e-14	0.83	Brain - Frontal Cortex (BA9)

ENSG00000185829.17	ARL17A	4.2e-14	0.64	Heart - Left Ventricle
ENSG00000185829.17	ARL17A	4.4e-14	0.66	Heart - Atrial Appendage
ENSG00000261575.2	RP11-259G18.1	6.5e-14	0.81	Prostate
ENSG00000262539.1	RP11-259G18.3	7.3e-14	1.1	Brain - Substantia nigra
ENSG00000185829.17	ARL17A	8.0e-14	0.65	Colon - Sigmoid
ENSG00000225190.10	PLEKHM1	9.2e-14	0.44	Brain - Cortex
ENSG00000176681.14	LRRC37A	1.1e-13	0.78	Brain - Cortex
ENSG00000261575.2	RP11-259G18.1	1.1e-13	0.64	Heart - Atrial Appendage
ENSG00000279685.2	MAPT-IT1	1.2e-13	0.49	Testis
ENSG00000225190.10	PLEKHM1	1.5e-13	0.41	Brain - Nucleus accumbens (basal ganglia)
ENSG00000261575.2	RP11-259G18.1	1.8e-13	0.84	Brain - Anterior cingulate cortex (BA24)
ENSG00000238083.7	LRRC37A2	2.0e-13	0.88	Cells - EBV-transformed lymphocytes
ENSG00000264589.2	MAPT-AS1	2.0e-13	-0.37	Muscle - Skeletal
ENSG00000185829.17	ARL17A	2.2e-13	0.54	Breast - Mammary Tissue
ENSG00000108379.9	WNT3	3.0e-13	0.43	Nerve - Tibial
ENSG00000120088.14	CRHR1	3.9e-13	0.50	Nerve - Tibial
ENSG00000185829.17	ARL17A	4.6e-13	0.89	Brain - Anterior cingulate cortex (BA24)
ENSG00000261575.2	RP11-259G18.1	5.1e-13	0.48	Artery - Tibial
ENSG00000264070.1	DND1P1	6.1e-13	0.76	Brain - Cerebellar Hemisphere
ENSG00000108379.9	WNT3	6.4e-13	0.36	Artery - Tibial
ENSG00000263503.1	MAPK8IP1P2	7.0e-13	1.1	Kidney - Cortex
ENSG00000108379.9	WNT3	7.1e-13	0.51	Esophagus - Gastroesophageal Junction
ENSG00000176681.14	LRRC37A	7.3e-13	0.89	Brain - Hypothalamus
ENSG00000238083.7	LRRC37A2	9.7e-13	1.0	Kidney - Cortex
ENSG00000185829.17	ARL17A	9.9e-13	0.51	Nerve - Tibial
ENSG00000261575.2	RP11-259G18.1	1.1e-12	0.57	Heart - Left Ventricle
ENSG00000261575.2	RP11-259G18.1	1.6e-12	0.60	Colon - Transverse
ENSG00000176681.14	LRRC37A	1.7e-12	0.70	Uterus
ENSG00000185829.17	ARL17A	2.0e-12	0.44	Skin - Sun Exposed (Lower leg)
ENSG00000185829.17	ARL17A	2.2e-12	0.78	Brain - Putamen (basal ganglia)
ENSG00000176681.14	LRRC37A	2.2e-12	0.75	Prostate
ENSG00000236234.1	AC091132.1	2.6e-12	-0.63	Testis
ENSG00000159314.11	ARHGAP27	2.6e-12	-0.52	Cells - EBV-transformed lymphocytes
ENSG00000263503.1	MAPK8IP1P2	2.7e-12	1.1	Brain - Substantia nigra
ENSG00000176681.14	LRRC37A	3.0e-12	0.73	Pituitary
ENSG00000261575.2	RP11-259G18.1	4.0e-12	0.82	Brain - Amygdala
ENSG00000185829.17	ARL17A	4.4e-12	0.60	Testis
ENSG00000232300.1	FAM215B	5.0e-12	0.52	Esophagus - Muscularis
ENSG00000232300.1	FAM215B	5.2e-12	0.48	Nerve - Tibial
ENSG00000236234.1	AC091132.1	5.9e-12	-0.26	Esophagus - Mucosa
ENSG00000280022.1	RP11-707O23.1	6.3e-12	0.88	Cells - EBV-transformed lymphocytes
ENSG00000108379.9	WNT3	6.8e-12	0.61	Adrenal Gland

ENSG00000120071.13	KANSL1	8.9e-12	0.22	Whole Blood
ENSG00000232300.1	FAM215B	1.4e-11	0.49	Lung
ENSG00000108379.9	WNT3	2.0e-11	0.38	Adipose - Subcutaneous
ENSG00000185829.17	ARL17A	3.1e-11	0.44	Skin - Not Sun Exposed (Suprapubic)
ENSG00000225190.10	PLEKHM1	3.6e-11	0.42	Adrenal Gland
ENSG00000185829.17	ARL17A	3.9e-11	0.44	Artery - Aorta
ENSG00000261575.2	RP11-259G18.1	4.5e-11	0.47	Adipose - Visceral (Omentum)
ENSG00000185829.17	ARL17A	4.7e-11	0.69	Prostate
ENSG00000176681.14	LRRC37A	5.1e-11	0.77	Brain - Frontal Cortex (BA9)
ENSG00000262500.1	MAPK8IP1P1	6.5e-11	1.0	Kidney - Cortex
ENSG00000261575.2	RP11-259G18.1	7.2e-11	0.85	Ovary
ENSG00000280022.1	RP11-707O23.1	7.7e-11	0.84	Brain - Spinal cord (cervical c-1)
ENSG00000236234.1	AC091132.1	9.4e-11	-0.61	Brain - Cerebellum
ENSG00000185829.17	ARL17A	1.0e-10	0.40	Artery - Tibial
ENSG00000280022.1	RP11-707O23.1	1.1e-10	0.71	Spleen
ENSG00000261575.2	RP11-259G18.1	1.1e-10	0.71	Small Intestine - Terminal Ileum
ENSG00000261575.2	RP11-259G18.1	1.2e-10	0.86	Uterus
ENSG00000266918.1	RP11-798G7.8	1.5e-10	0.46	Nerve - Tibial
ENSG00000264070.1	DND1P1	2.0e-10	1.1	Kidney - Cortex
ENSG00000108379.9	WNT3	2.0e-10	0.48	Colon - Sigmoid
ENSG00000264589.2	MAPT-AS1	2.1e-10	-0.60	Brain - Caudate (basal ganglia)
ENSG00000261575.2	RP11-259G18.1	2.1e-10	0.77	Vagina
ENSG00000204652.6	RPS26P8	2.1e-10	0.24	Testis
ENSG00000262539.1	RP11-259G18.3	2.3e-10	1.0	Kidney - Cortex
ENSG00000232300.1	FAM215B	2.9e-10	0.61	Pituitary
ENSG00000120071.13	KANSL1	3.5e-10	0.67	Cells - EBV-transformed lymphocytes
ENSG00000185829.17	ARL17A	3.8e-10	0.68	Brain - Hippocampus
ENSG00000261575.2	RP11-259G18.1	3.8e-10	0.46	Breast - Mammary Tissue
ENSG00000279685.2	MAPT-IT1	4.4e-10	-0.43	Liver
ENSG00000176681.14	LRRC37A	4.8e-10	0.72	Brain - Caudate (basal ganglia)
ENSG00000159314.11	ARHGAP27	4.9e-10	-0.34	Testis
ENSG00000267121.5	CTD-2020K17.1	5.4e-10	-0.55	Brain - Cerebellum
ENSG00000185829.17	ARL17A	6.2e-10	0.82	Brain - Amygdala
ENSG00000108379.9	WNT3	6.9e-10	0.49	Heart - Left Ventricle
ENSG00000176681.14	LRRC37A	7.2e-10	0.91	Brain - Spinal cord (cervical c-1)
ENSG00000120071.13	KANSL1	7.6e-10	0.29	Adipose - Subcutaneous
ENSG00000267198.1	RP11-798G7.6	9.9e-10	-0.49	Esophagus - Mucosa
ENSG00000225190.10	PLEKHM1	1.0e-9	-0.35	Testis
ENSG00000204650.14	LINC02210	1.1e-9	0.31	Brain - Cerebellar Hemisphere
ENSG00000120088.14	CRHR1	1.1e-9	0.38	Esophagus - Muscularis
ENSG00000073969.18	NSF	1.1e-9	-0.16	Esophagus - Mucosa
ENSG00000185829.17	ARL17A	1.2e-9	0.72	Ovary

ENSG00000225190.10	PLEKHM1	1.5e-9	0.16	Esophagus - Muscularis
ENSG00000266918.1	RP11-798G7.8	1.6e-9	0.40	Artery - Tibial
ENSG00000176681.14	LRRC37A	1.7e-9	0.49	Heart - Left Ventricle
ENSG00000186868.15	MAPT	2.1e-9	0.29	Skin - Not Sun Exposed (Suprapubic)
ENSG00000264589.2	MAPT-AS1	2.4e-9	-0.65	Brain - Spinal cord (cervical c-1)
ENSG00000108379.9	WNT3	2.6e-9	0.44	Colon - Transverse
ENSG00000185829.17	ARL17A	2.7e-9	0.65	Liver
ENSG00000176681.14	LRRC37A	3.2e-9	0.72	Brain - Hippocampus
ENSG00000186868.15	MAPT	3.7e-9	-0.32	Artery - Aorta
ENSG00000073969.18	NSF	3.8e-9	0.22	Testis
ENSG00000159314.11	ARHGAP27	4.8e-9	0.24	Brain - Caudate (basal ganglia)
ENSG00000176681.14	LRRC37A	4.8e-9	0.64	Ovary
ENSG00000185829.17	ARL17A	5.4e-9	0.60	Artery - Coronary
ENSG00000266918.1	RP11-798G7.8	5.5e-9	0.39	Adipose - Subcutaneous
ENSG00000108379.9	WNT3	6.1e-9	0.52	Artery - Coronary
ENSG00000266918.1	RP11-798G7.8	6.4e-9	0.57	Brain - Cerebellum
ENSG00000261575.2	RP11-259G18.1	6.8e-9	0.71	Brain - Hypothalamus
ENSG00000176681.14	LRRC37A	7.6e-9	0.65	Brain - Nucleus accumbens (basal ganglia)
ENSG00000279685.2	MAPT-IT1	7.8e-9	0.50	Brain - Cerebellum
ENSG00000204650.14	LINC02210	1.1e-8	0.37	Brain - Cerebellum
ENSG00000266918.1	RP11-798G7.8	1.3e-8	0.49	Colon - Sigmoid
ENSG00000225190.10	PLEKHM1	1.3e-8	0.17	Adipose - Subcutaneous
ENSG00000225190.10	PLEKHM1	1.9e-8	0.11	Cells - Cultured fibroblasts
ENSG00000176681.14	LRRC37A	2.0e-8	0.91	Brain - Substantia nigra
ENSG00000185829.17	ARL17A	2.2e-8	0.46	Esophagus - Gastroesophageal Junction
ENSG00000280022.1	RP11-707O23.1	2.6e-8	-0.52	Brain - Cerebellar Hemisphere
ENSG00000267344.1	CTB-39G8.3	2.9e-8	0.24	Skin - Not Sun Exposed (Suprapubic)
ENSG00000185829.17	ARL17A	3.2e-8	0.69	Spleen
ENSG00000185829.17	ARL17A	3.6e-8	0.70	Uterus
ENSG00000225190.10	PLEKHM1	3.7e-8	0.18	Lung
ENSG00000185829.17	ARL17A	4.2e-8	0.77	Minor Salivary Gland
ENSG00000073969.18	NSF	4.4e-8	0.19	Cells - Cultured fibroblasts
ENSG00000261575.2	RP11-259G18.1	4.4e-8	0.36	Lung
ENSG00000120071.13	KANSL1	4.5e-8	0.37	Colon - Sigmoid
ENSG00000176681.14	LRRC37A	4.7e-8	0.40	Esophagus - Mucosa
ENSG00000225190.10	PLEKHM1	4.7e-8	0.31	Brain - Caudate (basal ganglia)
ENSG00000262372.1	RP11-669E14.6	5.1e-8	-0.39	Testis
ENSG00000261575.2	RP11-259G18.1	5.4e-8	0.72	Cells - EBV-transformed lymphocytes
ENSG00000266918.1	RP11-798G7.8	6.1e-8	0.49	Colon - Transverse
ENSG00000225190.10	PLEKHM1	6.2e-8	0.39	Brain - Anterior cingulate cortex (BA24)
ENSG00000266918.1	RP11-798G7.8	6.7e-8	0.73	Uterus
ENSG00000108379.9	WNT3	6.8e-8	0.31	Testis

ENSG00000266918.1	RP11-798G7.8	7.3e-8	0.47	Esophagus - Gastroesophageal Junction
ENSG00000108379.9	WNT3	7.8e-8	0.40	Heart - Atrial Appendage
ENSG00000261575.2	RP11-259G18.1	7.9e-8	0.24	Muscle - Skeletal
ENSG00000176681.14	LRRC37A	8.6e-8	0.68	Brain - Putamen (basal ganglia)
ENSG00000176681.14	LRRC37A	1.2e-7	0.67	Brain - Anterior cingulate cortex (BA24)
ENSG00000176681.14	LRRC37A	1.3e-7	0.47	Stomach
ENSG00000236234.1	AC091132.1	1.3e-7	-0.56	Brain - Cerebellar Hemisphere
ENSG00000225190.10	PLEKHM1	1.3e-7	0.19	Breast - Mammary Tissue
ENSG00000120071.13	KANSL1	1.7e-7	0.36	Colon - Transverse
ENSG00000120071.13	KANSL1	1.8e-7	0.23	Adipose - Visceral (Omentum)
ENSG00000267344.1	CTB-39G8.3	1.8e-7	-0.17	Esophagus - Mucosa
ENSG00000108379.9	WNT3	2.0e-7	0.33	Stomach
ENSG00000225190.10	PLEKHM1	2.3e-7	0.34	Brain - Putamen (basal ganglia)
ENSG00000186868.15	MAPT	2.3e-7	-0.17	Colon - Transverse
ENSG00000073969.18	NSF	2.4e-7	-0.19	Esophagus - Muscularis
ENSG00000120088.14	CRHR1	2.4e-7	0.25	Muscle - Skeletal
ENSG00000266918.1	RP11-798G7.8	2.6e-7	0.38	Adipose - Visceral (Omentum)
ENSG00000108379.9	WNT3	2.7e-7	0.30	Breast - Mammary Tissue
ENSG00000264589.2	MAPT-AS1	3.0e-7	-0.62	Brain - Putamen (basal ganglia)
ENSG00000232300.1	FAM215B	3.3e-7	0.59	Prostate
ENSG00000185829.17	ARL17A	3.3e-7	0.54	Small Intestine - Terminal Ileum
ENSG00000176681.14	LRRC37A	4.9e-7	0.46	Colon - Transverse
ENSG00000176681.14	LRRC37A	5.1e-7	0.56	Vagina
ENSG00000120071.13	KANSL1	5.4e-7	0.24	Artery - Tibial
ENSG00000176681.14	LRRC37A	6.3e-7	0.68	Brain - Amygdala
ENSG00000266918.1	RP11-798G7.8	6.3e-7	0.39	Artery - Aorta
ENSG00000186868.15	MAPT	6.8e-7	-0.19	Artery - Tibial
ENSG00000261575.2	RP11-259G18.1	7.0e-7	0.56	Brain - Hippocampus
ENSG00000186868.15	MAPT	7.9e-7	-0.29	Brain - Cerebellum
ENSG00000225190.10	PLEKHM1	8.0e-7	0.32	Brain - Frontal Cortex (BA9)
ENSG00000266918.1	RP11-798G7.8	8.1e-7	0.46	Stomach
ENSG00000073969.18	NSF	8.3e-7	-0.20	Heart - Atrial Appendage
ENSG00000204652.6	RPS26P8	9.4e-7	0.72	Ovary
ENSG00000267121.5	CTD-2020K17.1	9.8e-7	-0.44	Brain - Cerebellar Hemisphere
ENSG00000186868.15	MAPT	0.0000010	0.17	Adipose - Visceral (Omentum)
ENSG00000073969.18	NSF	0.0000012	-0.17	Artery - Tibial
ENSG00000073969.18	NSF	0.0000013	-0.17	Artery - Aorta
ENSG00000228696.8	ARL17B	0.0000014	0.36	Skin - Not Sun Exposed (Suprapubic)
ENSG00000176681.14	LRRC37A	0.0000014	0.55	Adrenal Gland
ENSG00000176681.14	LRRC37A	0.0000015	0.62	Small Intestine - Terminal Ileum
ENSG00000176681.14	LRRC37A	0.0000018	0.54	Spleen
ENSG00000073969.18	NSF	0.0000025	-0.25	Ovary

ENSG00000264589.2	MAPT-AS1	0.0000025	-0.23	Brain - Cerebellar Hemisphere
ENSG00000136448.11	NMT1	0.0000026	-0.14	Cells - Cultured fibroblasts
ENSG00000159314.11	ARHGAP27	0.0000028	0.32	Brain - Cortex
ENSG00000225190.10	PLEKHM1	0.0000028	0.18	Adipose - Visceral (Omentum)
ENSG00000120071.13	KANSL1	0.0000028	0.28	Lung
ENSG00000266918.1	RP11-798G7.8	0.0000029	0.49	Pituitary
ENSG00000185829.17	ARL17A	0.0000031	0.28	Whole Blood
ENSG00000261575.2	RP11-259G18.1	0.0000033	0.50	Artery - Coronary
ENSG00000280022.1	RP11-707O23.1	0.0000045	-0.55	Brain - Hypothalamus
ENSG00000120088.14	CRHR1	0.0000045	0.37	Esophagus - Gastroesophageal Junction
ENSG00000136448.11	NMT1	0.0000045	-0.14	Thyroid
ENSG00000232300.1	FAM215B	0.0000046	0.41	Esophagus - Gastroesophageal Junction
ENSG00000267344.1	CTB-39G8.3	0.0000049	-0.39	Testis
ENSG00000232300.1	FAM215B	0.0000056	0.35	Esophagus - Mucosa
ENSG00000176681.14	LRRC37A	0.0000057	0.37	Heart - Atrial Appendage
ENSG00000261575.2	RP11-259G18.1	0.0000065	0.60	Minor Salivary Gland
ENSG00000266918.1	RP11-798G7.8	0.0000079	0.33	Skin - Not Sun Exposed (Suprapubic)
ENSG00000186868.15	MAPT	0.0000082	-0.28	Prostate
ENSG00000266918.1	RP11-798G7.8	0.0000085	0.33	Lung
ENSG00000236234.1	AC091132.1	0.0000090	-0.57	Cells - EBV-transformed lymphocytes
ENSG00000185294.6	SPPL2C	0.0000090	0.37	Brain - Frontal Cortex (BA9)
ENSG00000204650.14	LINC02210	0.0000098	-0.31	Liver
ENSG00000261575.2	RP11-259G18.1	0.000010	0.67	Brain - Substantia nigra
ENSG00000120071.13	KANSL1	0.000011	0.33	Brain - Cerebellum
ENSG00000186868.15	MAPT	0.000011	-0.21	Esophagus - Muscularis
ENSG00000186868.15	MAPT	0.000013	0.46	Spleen
ENSG00000120071.13	KANSL1	0.000013	0.35	Brain - Cerebellar Hemisphere
ENSG00000225190.10	PLEKHM1	0.000014	-0.066	Whole Blood
ENSG00000280022.1	RP11-707O23.1	0.000014	0.47	Brain - Cortex
ENSG00000228696.8	ARL17B	0.000014	0.31	Skin - Sun Exposed (Lower leg)
ENSG00000264589.2	MAPT-AS1	0.000014	-0.67	Brain - Substantia nigra
ENSG00000108379.9	WNT3	0.000015	0.24	Muscle - Skeletal
ENSG00000232300.1	FAM215B	0.000015	0.61	Ovary
ENSG00000265964.1	RP11-293E1.1	0.000015	-0.28	Testis
ENSG00000225190.10	PLEKHM1	0.000017	0.34	Brain - Amygdala
ENSG00000120071.13	KANSL1	0.000018	0.30	Small Intestine - Terminal Ileum
ENSG00000266918.1	RP11-798G7.8	0.000022	0.43	Pancreas
ENSG00000185829.17	ARL17A	0.000026	0.58	Cells - EBV-transformed lymphocytes
ENSG00000172992.11	DCAKD	0.000029	0.19	Artery - Tibial
ENSG00000261575.2	RP11-259G18.1	0.000030	0.63	Brain - Spinal cord (cervical c-1)
ENSG00000228696.8	ARL17B	0.000032	-0.27	Muscle - Skeletal
ENSG00000264589.2	MAPT-AS1	0.000032	-0.31	Brain - Hippocampus

ENSG00000266918.1	RP11-798G7.8	0.000035	0.28	Skin - Sun Exposed (Lower leg)
ENSG00000186868.15	MAPT	0.000035	0.12	Adipose - Subcutaneous
ENSG00000232300.1	FAM215B	0.000039	0.27	Skin - Sun Exposed (Lower leg)
ENSG00000267198.1	RP11-798G7.6	0.000041	-0.29	Skin - Sun Exposed (Lower leg)
ENSG00000232300.1	FAM215B	0.000042	0.34	Breast - Mammary Tissue
ENSG00000120071.13	KANSL1	0.000043	0.37	Pituitary
ENSG00000266918.1	RP11-798G7.8	0.000044	0.42	Brain - Cortex
ENSG00000266918.1	RP11-798G7.8	0.000049	0.37	Heart - Atrial Appendage
ENSG00000185829.17	ARL17A	0.000050	0.48	Vagina
ENSG00000186868.15	MAPT	0.000054	-0.19	Thyroid
ENSG00000266918.1	RP11-798G7.8	0.000054	0.56	Ovary
ENSG00000232300.1	FAM215B	0.000055	0.46	Artery - Coronary
ENSG00000172992.11	DCAKD	0.000057	0.16	Adipose - Subcutaneous
ENSG00000108379.9	WNT3	0.000063	0.24	Adipose - Visceral (Omentum)
ENSG00000120088.14	CRHR1	0.000066	-0.26	Brain - Caudate (basal ganglia)
ENSG00000267198.1	RP11-798G7.6	0.000067	-0.46	Brain - Cerebellum
ENSG00000228696.8	ARL17B	0.000071	-0.27	Whole Blood
ENSG00000120088.14	CRHR1	0.000074	-0.23	Brain - Hippocampus
ENSG00000136448.11	NMT1	0.000079	-0.23	Brain - Cerebellar Hemisphere
ENSG00000120088.14	CRHR1	0.000095	-0.26	Brain - Nucleus accumbens (basal ganglia)
ENSG00000172992.11	DCAKD	0.00011	0.19	Nerve - Tibial
ENSG00000159314.11	ARHGAP27	0.00012	0.12	Artery - Aorta
ENSG00000225190.10	PLEKHM1	0.00015	0.090	Artery - Tibial
ENSG00000266918.1	RP11-798G7.8	0.00017	0.28	Esophagus - Muscularis
ENSG00000260075.1	NSFP1	0.00017	-0.36	Esophagus - Gastroesophageal Junction
ENSG00000073969.18	NSF	0.00017	-0.15	Brain - Cerebellum
ENSG00000185294.6	SPPL2C	0.00020	0.11	Testis
ENSG00000120088.14	CRHR1	0.00021	-0.18	Thyroid
ENSG00000181513.14	ACBD4	0.00021	-0.12	Nerve - Tibial
ENSG00000172992.11	DCAKD	0.00027	0.16	Esophagus - Mucosa
ENSG00000131484.4	RP11-798G7.5	0.00027	-0.25	Thyroid

p-value, eQTL association p value; NES, normalized effect size.

Table S22. sQTL statistics for rs2106786 from GTEx v8 portal.

Gencode Id	Gene Symbol	Phenotype Id	Intron Id	P-Value	NES	Tissue
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13592:ENSG00000120071.13	46094701:46170855:clu_13592	1.1e-197	-1.8	Artery - Tibial
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_15040:ENSG00000120071.13	46094701:46170855:clu_15040	1.4e-193	-1.7	Skin - Sun Exposed (Lower leg)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12990:ENSG00000120071.13	46094701:46170855:clu_12990	6.9e-189	-1.6	Muscle - Skeletal
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14811:ENSG00000120071.13	46094701:46170855:clu_14811	1.7e-180	-1.7	Adipose - Subcutaneous
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14811:ENSG00000120071.13	46094701:46170855:clu_14811	1.7e-180	-1.7	Adipose - Subcutaneous
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_15568:ENSG00000120071.13	46094701:46170855:clu_15568	2.2e-177	-1.7	Thyroid
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14690:ENSG00000120071.13	46094701:46170855:clu_14690	3.0e-170	-1.7	Skin - Not Sun Exposed (Suprapubic)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_15235:ENSG00000120071.13	46094701:46170855:clu_15235	3.4e-160	-1.8	Nerve - Tibial
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11699:ENSG00000120071.13	46094701:46170855:clu_11699	4.3e-158	-1.6	Whole Blood
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14343:ENSG00000120071.13	46094701:46170855:clu_14343	1.5e-150	-1.7	Adipose - Visceral (Omentum)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_15467:ENSG00000120071.13	46094701:46170855:clu_15467	5.2e-144	-1.7	Lung
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13509:ENSG00000120071.13	46094701:46170855:clu_13509	2.3e-141	-1.6	Esophagus - Muscularis
ENSG00000186868.15	MAPT	chr17:45991586:45996399:clu_12984:ENSG00000186868.15	45991586:45996399:clu_12984	9.9e-141	-1.4	Muscle - Skeletal
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13547:ENSG00000120071.13	46094701:46170855:clu_13547	8.6e-135	-1.5	Esophagus - Mucosa
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14805:ENSG00000120071.13	46094701:46170855:clu_14805	1.7e-126	-1.8	Breast - Mammary Tissue
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14256:ENSG00000120071.13	46094701:46170855:clu_14256	4.5e-114	-1.7	Colon - Transverse
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_14256:ENSG00000120071.13	46094701:46170855:clu_14256	4.5e-114	-1.7	Colon - Transverse
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13119:ENSG00000120071.13	46094701:46170855:clu_13119	3.1e-113	-1.6	Artery - Aorta
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13238:ENSG00000120071.13	46094701:46170855:clu_13238	1.3e-110	-1.7	Colon - Sigmoid
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12921:ENSG00000120071.13	46094701:46170855:clu_12921	1.8e-109	-1.6	Esophagus - Gastroesophageal Junction
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12921:ENSG00000120071.13	46094701:46170855:clu_12921	1.8e-109	-1.6	Esophagus - Gastroesophageal Junction
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11609:ENSG00000120071.13	46094701:46170855:clu_11609	1.2e-93	-1.5	Heart - Left Ventricle
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13190:ENSG00000120071.13	46094701:46170855:clu_13190	8.1e-85	-1.5	Stomach
ENSG00000186868.15	MAPT	chr17:45985744:45987040:clu_25872:ENSG00000186868.15	45985744:45987040:clu_25872	6.3e-84	-1.5	Testis
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12954:ENSG00000120071.13	46094701:46170855:clu_12954	1.7e-83	-1.5	Heart - Atrial Appendage
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13961:ENSG00000120071.13	46094701:46170855:clu_13961	1.6e-76	-1.7	Pituitary
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12402:ENSG00000120071.13	46094701:46170855:clu_12402	1.3e-74	-1.6	Brain - Cerebellum
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_15027:ENSG00000204650.14	45630158:45645901:clu_15027	2.0e-72	1.1	Skin - Sun Exposed (Lower leg)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13420:ENSG00000120071.13	46094701:46170855:clu_13420	2.1e-70	-1.7	Prostate
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13420:ENSG00000120071.13	46094701:46170855:clu_13420	2.1e-70	-1.7	Prostate
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_14678:ENSG00000204650.14	45630158:45645901:clu_14678	6.1e-70	1.1	Skin - Not Sun Exposed (Suprapubic)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_25893:ENSG00000120071.13	46094701:46170855:clu_25893	6.0e-69	-1.4	Testis

ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_25826:ENSG00000204650.14	45630158:45645901:clu_25826	9.3e-65	1.3	Testis
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11713:ENSG00000120071.13	46094701:46170855:clu_11713	3.5e-64	-1.6	Adrenal Gland
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_15554:ENSG00000214425.7	45510293:45513185:clu_15554	3.3e-62	-1.2	Thyroid
ENSG00000266504.1	RP11-798G7.4	chr17:45510293:45513185:clu_15554:ENSG00000266504.1	45510293:45513185:clu_15554	3.3e-62	-1.2	Thyroid
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12290:ENSG00000120071.13	46094701:46170855:clu_12290	5.2e-61	-1.6	Artery - Coronary
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11789:ENSG00000120071.13	46094701:46170855:clu_11789	3.7e-60	-1.5	Pancreas
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_12631:ENSG00000120071.13	46094701:46170855:clu_12631	1.3e-56	-1.7	Spleen
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11885:ENSG00000120071.13	46094701:46170855:clu_11885	1.3e-56	-1.7	Brain - Cerebellar Hemisphere
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11869:ENSG00000120071.13	46094701:46170855:clu_11869	3.2e-54	-1.1	Cells - Cultured fibroblasts
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_15168:ENSG00000214425.7	45545676:45584572:clu_15168	1.8e-53	-1.2	Nerve - Tibial
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11793:ENSG00000120071.13	46094701:46170855:clu_11793	6.6e-51	-1.5	Brain - Cortex
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11869:ENSG00000120071.13	46094701:46170855:clu_11869	3.5e-48	-1.9	Minor Salivary Gland
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11840:ENSG00000120071.13	46094701:46170855:clu_11840	6.2e-48	-1.7	Ovary
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_13534:ENSG00000204650.14	45630158:45636283:clu_13534	5.1e-47	-1.0	Esophagus - Mucosa
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11396:ENSG00000120071.13	46094701:46170855:clu_11396	6.5e-46	-1.5	Brain - Caudate (basal ganglia)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_13074:ENSG00000120071.13	46094701:46170855:clu_13074	7.7e-46	-1.6	Small Intestine - Terminal Ileum
ENSG00000214425.7	LRRC37A4P	chr17:45482525:45490652:clu_12383:ENSG00000214425.7	45482525:45490652:clu_12383	1.1e-45	1.2	Brain - Cerebellum
ENSG00000225190.10	PLEKHM1	chr17:45482525:45490652:clu_12383:ENSG00000225190.10	45482525:45490652:clu_12383	1.1e-45	1.2	Brain - Cerebellum
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11663:ENSG00000120071.13	46094701:46170855:clu_11663	4.6e-45	-1.5	Brain - Nucleus accumbens (basal ganglia)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10441:ENSG00000120071.13	46094701:46170855:clu_10441	6.8e-45	-1.6	Cells - EBV-transformed lymphocytes
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_15024:ENSG00000214425.7	45510293:45513185:clu_15024	8.1e-43	-0.84	Skin - Sun Exposed (Lower leg)
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_15454:ENSG00000214425.7	45510293:45513185:clu_15454	1.3e-42	-1.0	Lung
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_12976:ENSG00000214425.7	45545676:45584572:clu_12976	7.0e-41	-0.94	Muscle - Skeletal
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11124:ENSG00000120071.13	46094701:46170855:clu_11124	1.7e-40	-1.6	Uterus
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10194:ENSG00000120071.13	46094701:46170855:clu_10194	1.5e-38	-1.4	Brain - Putamen (basal ganglia)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10194:ENSG00000120071.13	46094701:46170855:clu_10194	1.5e-38	-1.4	Brain - Putamen (basal ganglia)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11684:ENSG00000120071.13	46094701:46170855:clu_11684	1.9e-38	-1.6	Vagina
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10943:ENSG00000120071.13	46094701:46170855:clu_10943	3.2e-37	-1.4	Brain - Frontal Cortex (BA9)
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_13578:ENSG00000214425.7	45545676:45584572:clu_13578	1.0e-36	-0.98	Artery - Tibial
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_13498:ENSG00000204650.14	45630158:45645901:clu_13498	4.0e-34	0.87	Esophagus - Muscularis
ENSG00000214425.7	LRRC37A4P	chr17:45482525:45490652:clu_11865:ENSG00000214425.7	45482525:45490652:clu_11865	4.2e-34	1.2	Brain - Cerebellar Hemisphere
ENSG00000214425.7	LRRC37A4P	chr17:45482525:45490652:clu_11865:ENSG00000214425.7	45482525:45490652:clu_11865	4.2e-34	1.2	Brain - Cerebellar Hemisphere
ENSG00000225190.10	PLEKHM1	chr17:45482525:45490652:clu_11865:ENSG00000225190.10	45482525:45490652:clu_11865	4.2e-34	1.2	Brain - Cerebellar Hemisphere

ENSG00000214425.7	LRRC37A4P	chr17:45551537:45584572:clu_11854:ENSG00000214425.7	45551537:45584572:clu_11854	2.7e-33	-0.94	Cells - Cultured fibroblasts
ENSG00000214425.7	LRRC37A4P	chr17:45550732:45551440:clu_13531:ENSG00000214425.7	45550732:45551440:clu_13531	5.7e-33	-0.88	Esophagus - Mucosa
ENSG00000266504.1	RP11-798G7.4	chr17:45550732:45551440:clu_13531:ENSG00000266504.1	45550732:45551440:clu_13531	5.7e-33	-0.88	Esophagus - Mucosa
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11857:ENSG00000204650.14	45630158:45636283:clu_11857	7.4e-33	-0.94	Cells - Cultured fibroblasts
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_11870:ENSG00000204650.14	45630158:45645901:clu_11870	1.3e-32	1.2	Brain - Cerebellar Hemisphere
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_14675:ENSG00000214425.7	45510293:45513185:clu_14675	2.8e-32	-0.80	Skin - Not Sun Exposed (Suprapubic)
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_14675:ENSG00000214425.7	45510293:45513185:clu_14675	2.8e-32	-0.80	Skin - Not Sun Exposed (Suprapubic)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_11097:ENSG00000120071.13	46094701:46170855:clu_11097	3.1e-32	-1.4	Brain - Hypothalamus
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_15557:ENSG00000204650.14	45630158:45636283:clu_15557	2.2e-31	-0.82	Thyroid
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10078:ENSG00000120071.13	46094701:46170855:clu_10078	3.8e-31	-1.4	Brain - Anterior cingulate cortex (BA24)
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_14332:ENSG00000214425.7	45510293:45513185:clu_14332	4.0e-31	-0.91	Adipose - Visceral (Omentum)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10089:ENSG00000120071.13	46094701:46170855:clu_10089	3.0e-29	-1.3	Liver
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_9945:ENSG00000120071.13	46094701:46170855:clu_9945	3.3e-29	-1.4	Brain - Spinal cord (cervical c-1)
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_12906:ENSG00000214425.7	45545676:45584572:clu_12906	3.7e-29	-1.1	Esophagus - Gastroesophageal Junction
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_13945:ENSG00000204650.14	45630158:45636283:clu_13945	8.2e-29	-1.0	Pituitary
ENSG00000214425.7	LRRC37A4P	chr17:45510293:45513185:clu_13105:ENSG00000214425.7	45510293:45513185:clu_13105	8.4e-28	-0.92	Artery - Aorta
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_10387:ENSG00000120071.13	46094701:46170855:clu_10387	1.0e-27	-1.3	Brain - Hippocampus
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_9365:ENSG00000120071.13	46094701:46170855:clu_9365	1.4e-25	-1.3	Brain - Amygdala
ENSG00000186868.15	MAPT	chr17:45974471:45978375:clu_13537:ENSG00000186868.15	45974471:45978375:clu_13537	1.5e-25	-0.69	Esophagus - Mucosa
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_15014:ENSG00000225190.10	45458439:45475100:clu_15014	1.2e-24	-0.69	Skin - Sun Exposed (Lower leg)
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_12910:ENSG00000204650.14	45630158:45645901:clu_12910	1.5e-23	0.82	Esophagus - Gastroesophageal Junction
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_15224:ENSG00000204650.14	45630158:45636283:clu_15224	1.6e-23	-0.80	Nerve - Tibial
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_12388:ENSG00000120088.14	45630158:45645901:clu_12388	9.5e-23	0.83	Brain - Cerebellum
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_12388:ENSG00000204650.14	45630158:45645901:clu_12388	9.5e-23	0.83	Brain - Cerebellum
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_13227:ENSG00000204650.14	45630158:45645901:clu_13227	1.4e-22	0.81	Colon - Sigmoid
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_10929:ENSG00000120088.14	45630158:45645901:clu_10929	1.7e-21	1.1	Brain - Frontal Cortex (BA9)
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_10929:ENSG00000204650.14	45630158:45645901:clu_10929	1.7e-21	1.1	Brain - Frontal Cortex (BA9)
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_25854:ENSG00000214425.7	45545676:45584572:clu_25854	3.3e-21	0.85	Testis
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_14335:ENSG00000204650.14	45630158:45636283:clu_14335	6.9e-21	-0.80	Adipose - Visceral (Omentum)
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_14793:ENSG00000225190.10	45458439:45475100:clu_14793	2.6e-20	-0.63	Adipose - Subcutaneous
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_13582:ENSG00000204650.14	45630158:45636283:clu_13582	3.1e-20	-0.71	Artery - Tibial
ENSG00000214425.7	LRRC37A4P	chr17:45551537:45584572:clu_13405:ENSG00000214425.7	45551537:45584572:clu_13405	2.5e-18	-0.98	Prostate
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_11685:ENSG00000214425.7	45545676:45584572:clu_11685	3.5e-18	0.62	Whole Blood

ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_15163:ENSG00000225190.10	45458439:45475100:clu_15163	3.7e-18	-0.63	Nerve - Tibial
ENSG00000186868.15	MAPT	chr17:45974471:45978375:clu_12394:ENSG00000186868.15	45974471:45978375:clu_12394	6.5e-18	0.93	Brain - Cerebellum
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11830:ENSG00000204650.14	45630158:45636283:clu_11830	8.4e-18	-1.3	Ovary
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_11779:ENSG00000120088.14	45630158:45645901:clu_11779	1.7e-17	0.86	Brain - Cortex
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_11779:ENSG00000204650.14	45630158:45645901:clu_11779	1.7e-17	0.86	Brain - Cortex
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_12941:ENSG00000214425.7	45545676:45584572:clu_12941	2.7e-17	-0.85	Heart - Atrial Appendage
ENSG00000238083.7	LRRC37A2	chr17:46487140:46517362:clu_13004:ENSG00000238083.7	46487140:46517362:clu_13004	6.6e-17	0.56	Muscle - Skeletal
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_14788:ENSG00000204650.14	45630158:45636283:clu_14788	7.1e-17	-0.77	Breast - Mammary Tissue
ENSG00000214425.7	LRRC37A4P	chr17:45551537:45584572:clu_13942:ENSG00000214425.7	45551537:45584572:clu_13942	8.0e-17	-0.89	Pituitary
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_11650:ENSG00000120088.14	45630158:45645901:clu_11650	4.5e-16	0.95	Brain - Nucleus accumbens (basal ganglia)
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_11650:ENSG00000204650.14	45630158:45645901:clu_11650	4.5e-16	0.95	Brain - Nucleus accumbens (basal ganglia)
ENSG00000120088.14	CRHR1	chr17:45630158:45636283:clu_11383:ENSG00000120088.14	45630158:45636283:clu_11383	1.4e-15	-0.92	Brain - Caudate (basal ganglia)
ENSG00000120088.14	CRHR1	chr17:45630158:45636283:clu_11383:ENSG00000120088.14	45630158:45636283:clu_11383	1.4e-15	-0.92	Brain - Caudate (basal ganglia)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11383:ENSG00000204650.14	45630158:45636283:clu_11383	1.4e-15	-0.92	Brain - Caudate (basal ganglia)
ENSG00000186868.15	MAPT	chr17:45978440:45987040:clu_13230:ENSG00000186868.15	45978440:45987040:clu_13230	1.7e-15	0.66	Colon - Sigmoid
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_12932:ENSG00000185829.17	46528859:46570759:clu_12932	2.3e-15	0.83	Esophagus - Gastroesophageal Junction
ENSG00000186868.15	MAPT	chr17:45987095:45991460:clu_11659:ENSG00000186868.15	45987095:45991460:clu_11659	4.1e-15	-0.57	Brain - Nucleus accumbens (basal ganglia)
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_8959:ENSG00000120071.13	46094701:46170855:clu_8959	6.8e-15	-1.2	Brain - Substantia nigra
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_14798:ENSG00000214425.7	45545676:45584572:clu_14798	7.6e-15	-0.60	Adipose - Subcutaneous
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_13407:ENSG00000204650.14	45630158:45645901:clu_13407	1.2e-14	0.77	Prostate
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_12900:ENSG00000225190.10	45458439:45475100:clu_12900	2.3e-14	-0.69	Esophagus - Gastroesophageal Junction
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_14245:ENSG00000204650.14	45630158:45636283:clu_14245	2.7e-14	-0.71	Colon - Transverse
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_11085:ENSG00000120088.14	45630158:45645901:clu_11085	3.0e-14	0.89	Brain - Hypothalamus
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_11085:ENSG00000204650.14	45630158:45645901:clu_11085	3.0e-14	0.89	Brain - Hypothalamus
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_15457:ENSG00000204650.14	45630158:45636283:clu_15457	5.0e-14	-0.63	Lung
ENSG00000120071.13	KANSL1	chr17:46094701:46170855:clu_8625:ENSG00000120071.13	46094701:46170855:clu_8625	5.7e-14	-1.3	Kidney - Cortex
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_15548:ENSG00000225190.10	45458439:45475100:clu_15548	1.3e-13	-0.52	Thyroid
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_11064:ENSG00000214425.7	45545676:45584572:clu_11064	1.4e-13	-1.2	Uterus
ENSG00000204650.14	LINC02210	chr17:45639520:45640389:clu_12979:ENSG00000204650.14	45639520:45640389:clu_12979	1.4e-13	-0.49	Muscle - Skeletal
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_15006:ENSG00000159314.11	45398047:45402714:clu_15006	1.9e-13	-0.53	Skin - Sun Exposed (Lower leg)
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_15006:ENSG00000159314.11	45398047:45402714:clu_15006	1.9e-13	-0.53	Skin - Sun Exposed (Lower leg)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_12944:ENSG00000204650.14	45630158:45636283:clu_12944	2.0e-13	-0.67	Heart - Atrial Appendage

ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_13179:ENSG00000204650.14	45630158:45636283:clu_13179	2.3e-13	-0.72	Stomach
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_10067:ENSG00000204650.14	45630158:45636283:clu_10067	4.7e-13	-0.94	Brain - Anterior cingulate cortex (BA24)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_9935:ENSG00000204650.14	45630158:45636283:clu_9935	6.1e-13	-1.0	Brain - Spinal cord (cervical c-1)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_14801:ENSG00000204650.14	45630158:45636283:clu_14801	6.2e-13	-0.55	Adipose - Subcutaneous
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_12973:ENSG00000225190.10	45458439:45475100:clu_12973	7.2e-13	-0.42	Muscle - Skeletal
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_11680:ENSG00000225190.10	45458439:45475100:clu_11680	8.4e-13	-0.34	Whole Blood
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_11827:ENSG00000214425.7	45545676:45584572:clu_11827	2.0e-12	-1.1	Ovary
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11703:ENSG00000204650.14	45630158:45636283:clu_11703	2.3e-12	-0.86	Adrenal Gland
ENSG00000120088.14	CRHR1	chr17:45630158:45636283:clu_10184:ENSG00000120088.14	45630158:45636283:clu_10184	3.9e-12	-0.81	Brain - Putamen (basal ganglia)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_10184:ENSG00000204650.14	45630158:45636283:clu_10184	3.9e-12	-0.81	Brain - Putamen (basal ganglia)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_13108:ENSG00000204650.14	45630158:45636283:clu_13108	4.4e-12	-0.63	Artery - Aorta
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_14781:ENSG00000225190.10	45458439:45475100:clu_14781	4.6e-12	-0.63	Breast - Mammary Tissue
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_15246:ENSG00000185829.17	46528859:46570759:clu_15246	5.3e-12	0.59	Nerve - Tibial
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11111:ENSG00000204650.14	45630158:45636283:clu_11111	9.7e-12	-1.0	Uterus
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_10436:ENSG00000204650.14	45630158:45636283:clu_10436	1.0e-11	-1.0	Cells - EBV-transformed lymphocytes
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_11672:ENSG00000204650.14	45630158:45645901:clu_11672	1.6e-11	0.85	Vagina
ENSG00000186868.15	MAPT	chr17:45987095:45991460:clu_11391:ENSG00000186868.15	45987095:45991460:clu_11391	1.6e-11	-0.66	Brain - Caudate (basal ganglia)
ENSG00000225190.10	PLEKHM1	chr17:45482525:45490652:clu_25849:ENSG00000225190.10	45482525:45490652:clu_25849	1.7e-11	0.64	Testis
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_13489:ENSG00000225190.10	45458439:45475100:clu_13489	3.4e-11	-0.46	Esophagus - Muscularis
ENSG00000186868.15	MAPT	chr17:45971945:45974385:clu_14682:ENSG00000186868.15	45971945:45974385:clu_14682	3.4e-11	-0.48	Skin - Not Sun Exposed (Suprapubic)
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_13572:ENSG00000225190.10	45458439:45475100:clu_13572	4.0e-11	-0.45	Artery - Tibial
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_15480:ENSG00000185829.17	46528859:46570759:clu_15480	4.2e-11	0.57	Lung
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_14785:ENSG00000214425.7	45545676:45584572:clu_14785	8.3e-11	-0.62	Breast - Mammary Tissue
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_14827:ENSG00000185829.17	46528859:46570759:clu_14827	1.3e-10	0.51	Adipose - Subcutaneous
ENSG00000186868.15	MAPT	chr17:45971945:45974385:clu_15031:ENSG00000186868.15	45971945:45974385:clu_15031	1.4e-10	-0.45	Skin - Sun Exposed (Lower leg)
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_14658:ENSG00000159314.11	45398047:45402714:clu_14658	1.6e-10	-0.51	Skin - Not Sun Exposed (Suprapubic)
ENSG00000186868.15	MAPT	chr17:45974471:45978375:clu_13952:ENSG00000186868.15	45974471:45978375:clu_13952	2.4e-10	-0.67	Pituitary
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_15580:ENSG00000185829.17	46528859:46570759:clu_15580	2.6e-10	0.51	Thyroid
ENSG00000238083.7	LRRC37A2	chr17:46487140:46511511:clu_11877:ENSG00000238083.7	46487140:46511511:clu_11877	3.7e-10	-0.55	Cells - Cultured fibroblasts
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_10376:ENSG00000120088.14	45630158:45645901:clu_10376	5.8e-10	0.76	Brain - Hippocampus
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_10376:ENSG00000120088.14	45630158:45645901:clu_10376	5.8e-10	0.76	Brain - Hippocampus
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_10376:ENSG00000204650.14	45630158:45645901:clu_10376	5.8e-10	0.76	Brain - Hippocampus
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_11132:ENSG00000185829.17	46528859:46570759:clu_11132	5.9e-10	0.96	Uterus

ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_11693:ENSG00000225190.10	45458439:45475100:clu_11693	8.7e-10	-0.60	Adrenal Gland
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_13101:ENSG00000225190.10	45458439:45475100:clu_13101	1.0e-9	-0.52	Artery - Aorta
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11859:ENSG00000204650.14	45630158:45636283:clu_11859	1.4e-9	-0.98	Minor Salivary Gland
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_14265:ENSG00000185829.17	46528859:46570759:clu_14265	1.6e-9	0.62	Colon - Transverse
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11689:ENSG00000204650.14	45630158:45636283:clu_11689	1.8e-9	-0.43	Whole Blood
ENSG00000186868.15	MAPT	chr17:45971945:45974385:clu_11787:ENSG00000186868.15	45971945:45974385:clu_11787	2.0e-9	0.70	Brain - Cortex
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_11700:ENSG00000214425.7	45545676:45584572:clu_11700	2.1e-9	-0.76	Adrenal Gland
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_15449:ENSG00000225190.10	45458439:45475100:clu_15449	2.3e-9	-0.42	Lung
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_9353:ENSG00000120088.14	45630158:45645901:clu_9353	2.3e-9	0.84	Brain - Amygdala
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_9353:ENSG00000204650.14	45630158:45645901:clu_9353	2.3e-9	0.84	Brain - Amygdala
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_11777:ENSG00000204650.14	45630158:45636283:clu_11777	3.5e-9	-0.61	Pancreas
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_11850:ENSG00000225190.10	45458439:45475100:clu_11850	4.0e-9	-0.45	Cells - Cultured fibroblasts
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13603:ENSG00000185829.17	46528859:46570759:clu_13603	6.5e-9	0.47	Artery - Tibial
ENSG00000238083.7	LRRC37A2	chr17:46487140:46511511:clu_10450:ENSG00000238083.7	46487140:46511511:clu_10450	6.9e-9	-0.97	Cells - EBV-transformed lymphocytes
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_15213:ENSG00000159314.11	45398047:45402714:clu_15213	9.3e-9	-0.48	Nerve - Tibial
ENSG00000186868.15	MAPT	chr17:45971945:45974385:clu_10937:ENSG00000186868.15	45971945:45974385:clu_10937	9.3e-9	0.57	Brain - Frontal Cortex (BA9)
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_14666:ENSG00000225190.10	45458439:45475100:clu_14666	1.1e-8	-0.41	Skin - Not Sun Exposed (Suprapubic)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_12279:ENSG00000204650.14	45630158:45636283:clu_12279	1.3e-8	-0.71	Artery - Coronary
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13248:ENSG00000185829.17	46528859:46570759:clu_13248	1.4e-8	0.61	Colon - Sigmoid
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_14811:ENSG00000185829.17	46528859:46570759:clu_14811	1.6e-8	0.56	Breast - Mammary Tissue
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13129:ENSG00000185829.17	46528859:46570759:clu_13129	1.7e-8	0.54	Artery - Aorta
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_12371:ENSG00000159314.11	45398047:45402714:clu_12371	2.2e-8	-0.65	Brain - Cerebellum
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_14232:ENSG00000225190.10	45458439:45475100:clu_14232	2.9e-8	-0.46	Colon - Transverse
ENSG00000186868.15	MAPT	chr17:45987095:45991460:clu_10074:ENSG00000186868.15	45987095:45991460:clu_10074	3.0e-8	-0.51	Brain - Anterior cingulate cortex (BA24)
ENSG00000186868.15	MAPT	chr17:45978440:45987040:clu_14248:ENSG00000186868.15	45978440:45987040:clu_14248	3.7e-8	0.49	Colon - Transverse
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_14355:ENSG00000185829.17	46528859:46570759:clu_14355	4.8e-8	0.47	Adipose - Visceral (Omentum)
ENSG00000186868.15	MAPT	chr17:45962470:45964397:clu_11877:ENSG00000186868.15	45962470:45964397:clu_11877	5.2e-8	-0.57	Brain - Cerebellar Hemisphere
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13199:ENSG00000185829.17	46528859:46570759:clu_13199	5.5e-8	0.59	Stomach
ENSG00000186868.15	MAPT	chr17:45987095:45991460:clu_12948:ENSG00000186868.15	45987095:45991460:clu_12948	7.1e-8	-0.56	Heart - Atrial Appendage
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_14328:ENSG00000225190.10	45458439:45475100:clu_14328	7.2e-8	-0.43	Adipose - Visceral (Omentum)
ENSG00000120088.14	CRHR1	chr17:45630158:45645901:clu_8948:ENSG00000120088.14	45630158:45645901:clu_8948	7.6e-8	0.99	Brain - Substantia nigra
ENSG00000204650.14	LINC02210	chr17:45630158:45645901:clu_8948:ENSG00000204650.14	45630158:45645901:clu_8948	7.6e-8	0.99	Brain - Substantia nigra
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_11854:ENSG00000159314.11	45398047:45402714:clu_11854	1.1e-7	-0.73	Brain - Cerebellar Hemisphere

ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_11379:ENSG00000214425.7	45545676:45584572:clu_11379	1.5e-7	-0.72	Brain - Caudate (basal ganglia)
ENSG00000238083.7	LRRC37A2	chr17:46487140:46517362:clu_13966:ENSG00000238083.7	46487140:46517362:clu_13966	1.5e-7	0.63	Pituitary
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13519:ENSG00000185829.17	46528859:46570759:clu_13519	1.6e-7	0.45	Esophagus - Muscularis
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_11078:ENSG00000225190.10	45458439:45475100:clu_11078	1.8e-7	-0.65	Brain - Hypothalamus
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_12639:ENSG00000185829.17	46528859:46570759:clu_12639	1.9e-7	0.75	Spleen
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_12414:ENSG00000185829.17	46528859:46570759:clu_12414	2.0e-7	0.65	Brain - Cerebellum
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_13175:ENSG00000214425.7	45545676:45584572:clu_13175	2.5e-7	0.54	Stomach
ENSG00000265547.1	RP11-293E1.2	chr17:45731944:45734284:clu_25866:ENSG00000265547.1	45731944:45734284:clu_25866	2.7e-7	-0.49	Testis
ENSG00000265964.1	RP11-293E1.1	chr17:45731944:45734284:clu_25866:ENSG00000265964.1	45731944:45734284:clu_25866	2.7e-7	-0.49	Testis
ENSG00000120088.14	CRHR1	chr17:45821440:45829215:clu_11872:ENSG00000120088.14	45821440:45829215:clu_11872	3.3e-7	-0.59	Brain - Cerebellar Hemisphere
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_13521:ENSG00000225190.10	45458439:45475100:clu_13521	3.8e-7	-0.32	Esophagus - Mucosa
ENSG00000225190.10	PLEKHM1	chr17:45458439:45475100:clu_13521:ENSG00000225190.10	45458439:45475100:clu_13521	3.8e-7	-0.32	Esophagus - Mucosa
ENSG00000159314.11	ARHGAP27	chr17:45401668:45402714:clu_25836:ENSG00000159314.11	45401668:45402714:clu_25836	4.6e-7	-0.51	Testis
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_14701:ENSG00000185829.17	46528859:46570759:clu_14701	5.0e-7	0.43	Skin - Not Sun Exposed (Suprapubic)
ENSG00000176681.14	LRRC37A	chr17:46331981:46332552:clu_25905:ENSG00000176681.14	46331981:46332552:clu_25905	5.6e-7	0.51	Testis
ENSG00000238083.7	LRRC37A2	chr17:46553449:46555152:clu_15053:ENSG00000238083.7	46553449:46555152:clu_15053	5.8e-7	-0.25	Skin - Sun Exposed (Lower leg)
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13429:ENSG00000185829.17	46528859:46570759:clu_13429	6.2e-7	0.63	Prostate
ENSG00000159314.11	ARHGAP27	chr17:45398047:45402714:clu_14770:ENSG00000159314.11	45398047:45402714:clu_14770	6.6e-7	-0.45	Breast - Mammary Tissue
ENSG00000185829.17	ARL17A	chr17:46528859:46570759:clu_13973:ENSG00000185829.17	46528859:46570759:clu_13973	0.0000012	0.60	Pituitary
ENSG00000159314.11	ARHGAP27	chr17:45397074:45397949:clu_13509:ENSG00000159314.11	45397074:45397949:clu_13509	0.0000018	0.39	Esophagus - Mucosa
ENSG00000159314.11	ARHGAP27	chr17:45397074:45397949:clu_11669:ENSG00000159314.11	45397074:45397949:clu_11669	0.0000020	0.33	Whole Blood
ENSG00000186868.15	MAPT	chr17:45987095:45991460:clu_14798:ENSG00000186868.15	45987095:45991460:clu_14798	0.0000024	-0.34	Breast - Mammary Tissue
ENSG00000186868.15	MAPT	chr17:45990075:45991460:clu_11603:ENSG00000186868.15	45990075:45991460:clu_11603	0.0000026	0.43	Heart - Left Ventricle
ENSG00000214425.7	LRRC37A4P	chr17:45545676:45584572:clu_14241:ENSG00000214425.7	45545676:45584572:clu_14241	0.0000044	0.46	Colon - Transverse
ENSG00000233175.2	CTD-2020K17.3	chr17:45240604:45241111:clu_11351:ENSG00000233175.2	45240604:45241111:clu_11351	0.0000052	0.55	Brain - Caudate (basal ganglia)
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_12620:ENSG00000204650.14	45630158:45636283:clu_12620	0.0000055	-0.60	Spleen
ENSG00000204650.14	LINC02210	chr17:45630158:45636283:clu_12620:ENSG00000204650.14	45630158:45636283:clu_12620	0.0000055	-0.60	Spleen
ENSG00000073969.18	NSF	chr17:46704854:46710963:clu_13008:ENSG00000073969.18	46704854:46710963:clu_13008	0.0000065	0.27	Muscle - Skeletal
ENSG00000238083.7	LRRC37A2	chr17:46517433:46520212:clu_13522:ENSG00000238083.7	46517433:46520212:clu_13522	0.0000067	-0.41	Esophagus - Muscularis

p-value, eQTL association p value; NES, normalized effect size.

Table S23. MR results using BAG as exposure.

outcome	effect	se	pval	padj	Model
AD	-0.0034	0.0062	0.59	0.74	IVW
BIP	0.0467	0.0276	9.04x10-2	0.15	IVW
MDD	-0.0063	0.0212	0.77	0.77	IVW
PD	-0.4012	0.1413	**4.54x10-3**	**2.27x10-2**	IVW
SCZ	-0.0854	0.0372	**2.18x10-2**	5.46x10-2	IVW
AD	-0.0542	0.0172	**1.96x10-2**	**4.9x10-2**	Egger
BIP	0.0765	0.1162	0.53	0.53	Egger
MDD	0.1092	0.0874	0.26	0.32	Egger
PD	-1.4666	0.4074	**1.14x10-2**	**4.9x10-2**	Egger
SCZ	-0.3272	0.1254	**4.02x10-2**	6.7x10-2	Egger
AD	-0.0098	0.0061	0.11	0.14	wMed
BIP	0.0661	0.0272	**1.51x10-2**	**3.77x10-2**	wMed
MDD	-0.0207	0.02485	0.41	0.41	wMed
PD	-0.2741	0.0865	**1.53x10-3**	**7.65x10-3**	wMed
SCZ	-0.0537	0.0308	8.11x10-2	0.14	wMed
AD	-0.0034	0.0062	0.61	0.76	PRESSO
BIP	0.0467	0.0276	0.13	0.22	PRESSO
MDD	-0.0063	0.0212	0.77	0.77	PRESSO
PD	-0.2371	0.0861	7.06x10-2	0.22	PRESSO
SCZ	-0.073	0.0344	8.72x10-2	0.22	PRESSO
AD	0.0018	0.0011	0.11	0.197	RAPS
BIP	0.0127	0.0051	**1.27x10-2**	**4.87x10-2**	RAPS
MDD	0.004	0.0042	0.34	0.43	RAPS
PD	-0.0237	0.0101	**1.95x10-2**	**4.88x1-0-2**	RAPS
SCZ	0.0017	0.0051	0.74	0.738	RAPS
AD	2.14729	1	0.98	0.98	CAUSE
BIP	0.8014	1	0.79	0.98	CAUSE
MDD	1.9085	1	0.97	0.98	CAUSE
PD	-0.1551	1	0.44	0.98	CAUSE
SCZ	0.4667	1	0.68	0.98	CAUSE

The six MR models are: IVW, Inverse variance weighted model; Egger, Egger regression; wMed, weighted median; PRESSO, MR-PRESSO; RAPS, MR-RAPS and CAUSE. causal effect size padj, FDR corrected p value; AD, Alzheimer's disease; BIP, bipolar disorder; MDD, major depression disorder; PD, Parkinson's disease; SCZ, schizophrenia. Note, negative CAUSE estimates indicate the existence of causal relations.

Table S24. MR results using disorders as exposure.

exposure	effect	se	pval	padj	Model
AD	0.69	0.21	**7.9x10-4**	**3.95x10-3**	IVW
BIP	0.15	0.06	**1.35x10-2**	**3.37x10-2**	IVW
MDD	0.32	0.17	6.48x10-2	8.1x10-2	IVW
PD	-0.25	0.12	3.35x10-2	5.59x10-2	IVW
SCZ	0.06	0.04	0.14	0.14	IVW
AD	0.48	0.30	0.13	0.16	Egger
BIP	0.68	0.22	**3.67x10-3**	**9.18x10-3**	Egger
MDD	0.29	0.91	0.76	0.76	Egger
PD	-1.11	0.31	**1.96x10-3**	**9.18x10-3**	Egger
SCZ	0.39	0.14	**5.66x10-3**	**9.44x10-3**	Egger
AD	0.67	0.22	**2.40x10-3**	**1.2x10-2**	wMed
BIP	0.14	0.08	9.71x10-2	0.18	wMed
MDD	0.33	0.23	0.16	0.19	wMed
PD	-0.02	0.07	0.75	0.75	wMed
SCZ	0.08	0.05	0.11	0.18	wMed
AD	0.69	0.21	**2.22x10-3**	**1.11x10-2**	PRESSO
BIP	0.15	0.06	**1.67x10-2**	**4.11x10-2**	PRESSO
MDD	0.32	0.14	5.87x10-2	0.07.34x10-2	PRESSO
PD	-0.05	0.06	0.36	0.36	PRESSO
SCZ	0.08	0.04	**2.46x10-2**	**04.11x10-2**	PRESSO
AD	0.50	0.11	**6.97x10-6**	**3.485x10-5**	RAPS
BIP	0.05	0.02	**9.05x10-3**	**0.01.8x10-2**	RAPS
MDD	0.02	0.03	0.56	0.56	RAPS
PD	-0.01	0.01	0.451	0.556	RAPS
SCZ	0.04	0.02	**1.08x10-2**	**1.8x10-2**	RAPS
AD	-0.59	1	0.28	0.99	CAUSE
BIP	1.01	1	0.84	0.99	CAUSE
MDD	3.18	1	0.99	0.99	CAUSE
PD	1.44	1	0.93	0.99	CAUSE
SCZ	-0.09	1	0.47	0.99	CAUSE

The six MR models are: IVW, Inverse variance weighted model; Egger, Egger regression; wMed, weighted median; PRESSO, MR-PRESSO; RAPS, MR-RAPS and CAUSE. causal effect size padj, FDR corrected p value; AD, Alzheimer's disease; BIP, bipolar disorder; MDD, major depression disorder; PD, Parkinson's disease; SCZ, schizophrenia. Note, negative CAUSE estimates indicate the existence of causal relations.

References

1. Consortium, S.W.G.o.t.P.G. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 421–427 (2014).
2. Ripke, S. Ricopili Pipeline And Standards of GWAS Analyses. *European Neuropsychopharmacology* **29**, S713-S714 (2019).
3. Lemieux Perreault, L.-P., Legault, M.-A., Asselin, G. & Dubé, M.-P. genipe: an automated genome-wide imputation pipeline with automatic reporting and statistical tools. *Bioinformatics* **32**, 3661-3663 (2016).
4. Rimol, L.M. *et al.* Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. *Biological psychiatry* **68**, 41-50 (2010).
5. Tønnesen, S. *et al.* White matter aberrations and age-related trajectories in patients with schizophrenia and bipolar disorder revealed by diffusion tensor imaging. *Scientific Reports* **8**, 14129 (2018).
6. Kelly, S. *et al.* Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Molecular Psychiatry* **23**, 1261-1269 (2018).
7. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
8. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature Genetics* **51**, 343-353 (2019).
9. Sul, J.H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively Identifying eQTLs from Multiple Tissues by Combining Mixed Model and Meta-analytic Approaches. *PLOS Genetics* **9**, e1003491 (2013).

III

Title

Characterizing personalized neuropathology in dementia and mild cognitive impairment with explainable artificial intelligence

Authors

Esten H. Leonardsen[1,2,*], Karin Persson[3,4], Edvard Grødem[1,5], Nicola Dinsdale[6], Till Schellhorn[7,8], James M. Roe[1], Didac Vidal-Piñeiro[1], Øystein Sørensen[1], Tobias Kaufmann[2,9,10], Eric Westman[11], Andre Marquand[12], Geir Selbæk[3,4], Ole A. Andreassen[2], Thomas Wolfers[1,2,9,†], Lars T. Westlye[1,2,13,†], Yunpeng Wang[1,†] for the Alzheimer's Disease Neuroimaging Initiative[‡] and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing[§]

Affiliations

1. Department of Psychology, University of Oslo, Oslo, Norway
2. Norwegian Centre for Mental Disorders Research (NORMENT), Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway
3. The Norwegian National Centre for Ageing and Health, Vestfold Hospital Trust, Norway
4. Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway
5. Computational Radiology & Artificial Intelligence (CRAI) Unit, Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway
6. Oxford Machine Learning in NeuroImaging (OMNI) Lab, University of Oxford, UK
7. Institute of Clinical Medicine, University of Oslo, Oslo, Norway
8. Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway
9. Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Germany
10. German Center for Mental Health (DZPG)
11. Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences, and Society, Karolinska Institutet, Stockholm, Sweden
12. Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, Netherlands
13. KG Jebsen Center for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway

† These authors contributed equally

‡ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

§ Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

* Corresponding author:
Esten H. Leonardsen,
Forskningsveien 3A, Harald Schjelderups hus, 0373 Oslo, Norway,
estenhl@uio.no

Abstract

Deep learning approaches for clinical predictions based on magnetic resonance imaging data have shown great promise as a translational technology for diagnosis and prognosis in neurological disorders, but its clinical impact has been limited. This is partially attributed to the opaqueness of deep learning models, causing insufficient understanding of what underlies their decisions. To overcome this, we trained convolutional neural networks on brain scans to differentiate dementia patients from healthy controls, and applied layerwise relevance propagation to procure individual-level explanations of the model predictions. Through extensive validations we demonstrate that deviations recognized by the model corroborate existing knowledge of neuropathology in dementia. By employing the explainable dementia classifier in a longitudinal dataset of patients with mild cognitive impairment, we show that the spatially rich explanations complement the model prediction when forecasting transition to dementia and help characterize the biological manifestation of disease in the individual brain. Overall, our work exemplifies the clinical potential of explainable artificial intelligence in precision medicine.

Introduction

Since its invention in the 1970s, magnetic resonance imaging (MRI) has provided an opportunity to non-invasively examine the body from the inside. In neuroscience, images acquired with MRI scanners have been used to identify how the brains of patients with various neurological disorders differ from their healthy counterparts. Stereotypically, this has been done by collecting data from a group of patients with a given disorder and a comparable group of healthy controls, on which traditional statistical inference is applied to identify spatial locations of the brain where the groups differ [1]. Typically, these locations are not atomic locations identified by a spatial coordinate, but rather morphological regions defined by an atlas, derived from empirical or theoretical insights of how the brain is structured. Differences between groups are described using morphometric properties like thickness or volume of these prespecified regions. A large benefit of this approach is the innate interpretability of the results: on average, patients with a given disorder deviate in a specific region of the brain in a comprehensible manner. Furthermore, the high degree of localization offered by modern brain scans allows for accurate characterization of where and how the brain of an individual deviates from an expected, typically healthy, norm [2]. However, the effects which are found are typically small [3] with limited

predictive power at the individual level [4,5], which in turn has raised questions about whether these analytical methods are expressive enough to model complex mental or clinical phenomena [6]. As an alternative, new conceptual approaches are proposed, advocating modelling frameworks with increased expressive power that allow for group differences through complex, non-linear interactions between multiple, potentially distant, parts of the brain [7], with a focus on prediction [8]. Such modelling flexibility is naturally achieved with artificial neural networks (ANNs), a class of statistical learning methods that combines aspects of data at multiple levels of abstraction, to accurately solve a predictive task [9]. However, while this often yields high predictive performance, e.g. by demonstrating clinically sufficient case-control classification accuracy for certain conditions, it comes at the cost of interpretation, as the models employ decision rules not trivially understandable by humans [10]. When the goal of the analysis is clinical, supporting the diagnosis and treatment of someone affected by a potential disorder, this opaqueness presents a substantial limitation. Thus, development and empirical validation of new methods within clinical neuroimaging that combine predictive efficacy with individual-level interpretability is imperative, to facilitate trust in how the system is working, and to accurately describe inter-individual heterogeneity.

With more than 55 million individuals afflicted worldwide [11], over 25 million disability-adjusted life years lost [12,13] and a cost exceeding one trillion USD yearly [14], dementia is a prime example of a neurological disorders that incur a monumental global burden. Due to the global aging population the prevalence is expected to nearly triple by 2050 [15], inciting a demand for technological solutions to facilitate handling the upcoming surge of patients. Dementia is a complex and progressive clinical condition [16] with multiple causal determinants and moderators. Alzheimer's disease (AD) is the most common form and accounts for 60%-80% of all cases [11]. However, the brain pathologies underlying different subtypes of dementia are not disjoint, but often co-occur [17–19], and have neuropathological commonalities [20]. The most prominent is neurodegeneration, occurring in both specific regions like the hippocampus, and globally across the brain [21], and inter-individual variations in the localization of atrophy has been associated with impairments in specific cognitive domains [22,23]. Thus, the biological manifestation of dementia in the brain is heterogeneous [24], resulting in distinctive cognitive and functional deficits [20], highlighting the need for precise and personalized approaches to diagnosis. For patients with

mild cognitive impairment (MCI), a potential clinical precursor to dementia, providing individualized characterizations of the underlying etiological disease at an early stage could widen the window for early interventions [25], alleviate uncertainty about the condition, and help with planning for the future [26].

In dementia, ANNs, and particularly convolutional neural networks (CNNs), have been applied to brain MRIs to differentiate patients from controls [27,28], prognosticate outcomes [29], and differentially diagnose subtypes [30]. However, while research utilizing this technology has been influential, clinical translations are scarce [31]. Where techniques for segmenting brain tumours or detecting lesions typically produce segmentation masks that are innately interpretable, predicting a complex diagnosis would entail compressing all information contained in a high-dimensional brain scan into a single number. Using deep learning, the decisions underlying this immense reduction are obfuscated, both from the developer of the system, the clinical personnel using it, and the patient ultimately impacted by the decision. This black box nature is broadly credited for the low levels of adoption in safety-critical domains like medicine [32]. Responding to this limitation, explainable artificial intelligence (XAI) provides methodology to explain the behaviour of ANNs [33]. The nature of these explanations varies, e.g. by what type of model is to be explained, what conceptual level the explanation is at, and who it is tailored for [34,35]. In computer vision, XAI typically aims for post-hoc explanations of individual decisions, explaining why a model arrived at a given prediction for a given image. Explanations are often provided in a visual format, as a heatmap indicating how different regions of the image contribute to the prediction [36]. Layerwise Relevance Propagation (LRP) is a variant of such a method, based on propagating relevance from the prediction-space, backwards through all layers of the model to the image-space, to form a relevance map [37]. A major advantage of LRP is its intuitive interpretation: by construction, the total amount of relevance which denotes contribution to the prediction is kept fixed between layers. Thus, the relevance propagated back to an input voxel is directly indicative of the influence of that exact voxel to the prediction. Recently, several studies have applied both LRP and other explainable AI methods to dementia [38], finding that the heatmaps generally highlight regions known to change in dementia [39,40]. However, the possibility of utilizing the fine-grained, individual, heatmaps produced by LRP to accurately

characterize individualized disease manifestations has not been explored, despite its potential for supporting clinical decisions towards precision medicine.

In the present study, we applied techniques from deep learning and XAI on MRI scans of the brain to make explainable and clinically relevant predictions for dementia at the individual level (Figure 1). Using a state-of-the-art architecture for neuroimaging data, we trained CNNs to differentiate patients diagnosed with dementia from healthy controls based on T1-weighted structural MRIs. We implemented LRP on top of the trained models to form a computational pipeline producing individual-level explanations in the form of relevance maps alongside the model predictions. The relevance maps were validated in a subset of dementia patients, both in a qualitative comparison with existing knowledge of the anatomical distribution of neuropathology, and in a quantitative, predictive context. Next, we applied the pipeline to a large, longitudinal dataset of MCI patients to create individual morphological records, a proposed data format for tracking and visualizing disease progression. Finally, we investigated the clinical utility of these records for stratifying patients, both in terms of their specific clinical profile, and progression of the disease. To facilitate reproducibility and improve the translational value of our work, the trained models and the complete explainable pipeline is made accessible on GitHub.

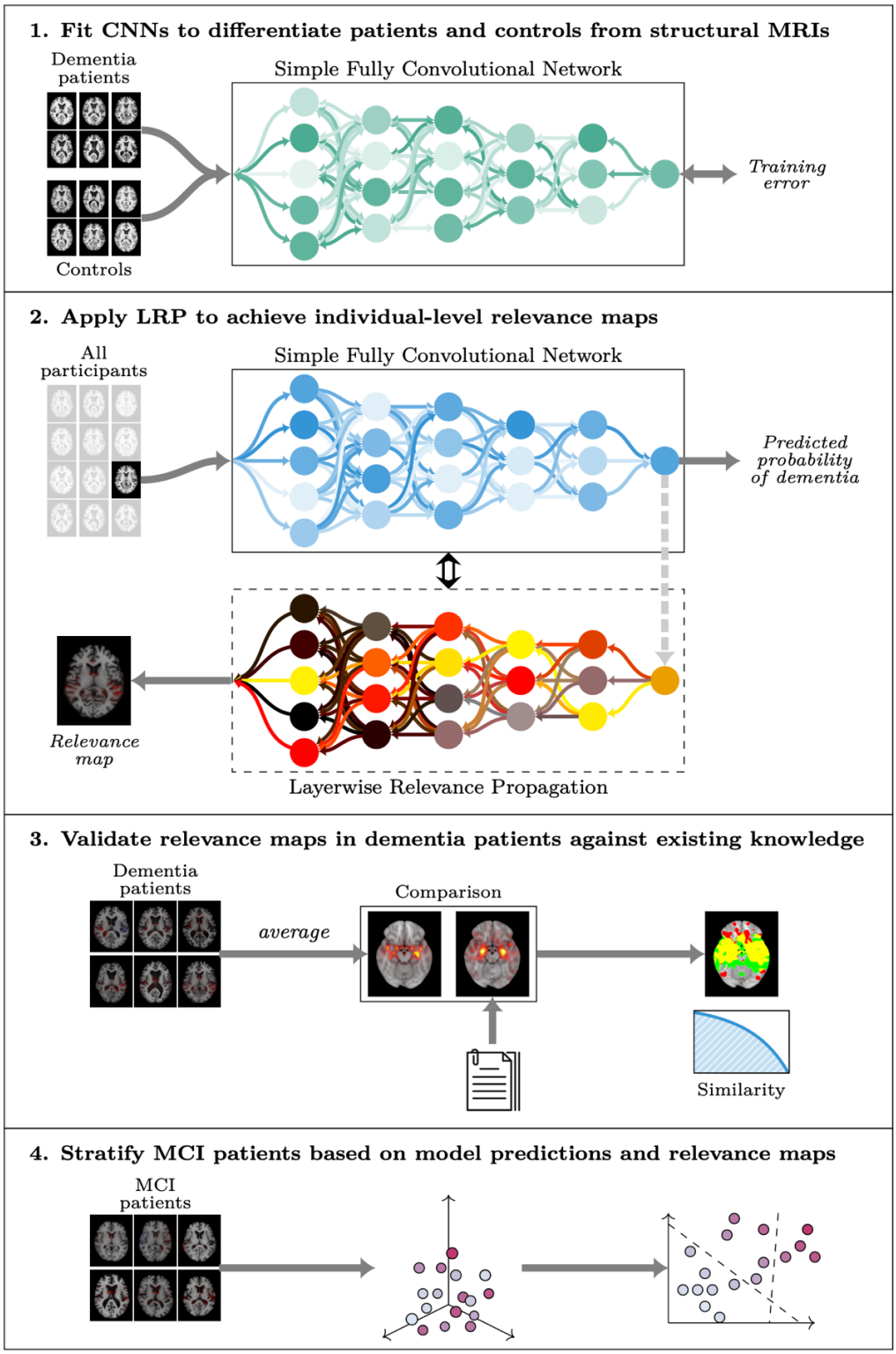


Figure 1: Overview of the modelling process. *The modelling process consisted of four disjoint steps. First, we fit multiple Simple Fully Convolutional Networks to classify dementia patients*

and healthy controls based on structural MRIs. Then we applied the best models to generate out-of-sample predictions and relevance maps for all participants. Next, we validated the relevance maps against existing knowledge using a meta-analysis to generate a statistical reference map. Finally, we employed the full pipeline in an exploratory analysis to stratify patients with mild cognitive impairment (MCI).

Results

We compiled MRI data from multiple sources (Supplementary Table 1) into a dataset of heterogeneous dementia patients (n=854, age range=47-95, 47% females, Table 1) based on various diagnoses (Probable AD, vascular dementia, other/unspecified dementia) and diagnostic criteria for inclusion (Supplementary Table 2), and a set of controls strictly matched on site, age, and sex of equal size. We trained multiple CNNs to differentiate between the groups, employing a nested validation approach utilizing all available timepoints for participants in three training folds and a single randomly selected timepoint for participants in separate validation and test folds. When stacking the out-of-sample predictions for all participants together (n=1708), for each fold using the model with the best validation performance, we observed satisfactory discrimination with a combined area under the receiver operating characteristics curve (AUC) of 0.908 (0.904-0.920 split across folds, Supplementary Figure 1), and an accuracy of 84.95% (83.04%-87.13%, Supplementary Table 3). This is slightly below with what is commonly achieved in similar studies classifying a specific subtype (typically AD) in a single dataset [28]. When grouping the out-of-sample predictions by site, each with a different acquisition protocol and diagnostic criteria, our AUCs ranged from 0.666 to 0.997 and accuracies from 58.33% to 97.36% (Supplementary Table 4). The best results were achieved in the datasets with a large number of datapoints, and a specific, clinical AD diagnosis, as opposed to those where we employed e.g. solely a mini-mental state examination (MMSE) threshold.

Cohort	Participants	Mean age (±std)	Sex (F/M)	Subdiagnoses (n)
Healthy controls	854	75.13±7.81	401/453	
Dementia patients	854	74.82±7.84	401/453	
MCI patients	1256	75.09±7.62	719/537	Improving (80), stable (754), progressive (304), other (118)
Total	**2964**	**75.08±7.65**	**1521/1443**	

Table 1: Cohorts. *Key characteristics of the cohorts used for training and testing the models, and further exploratory analyses.*

Relevance maps highlight predictive brain regions in individuals with dementia

Based on the classifiers with the highest AUCs in the validation sets, we built an explainable pipeline for dementia prediction, $LRP_{dementia}$, using composite LRP [41], and a strategy to prioritize regions of the brain that contributed positively towards a prediction of dementia (Supplementary Table 5). Using this pipeline, we computed out-of-sample relevance maps for all participants by applying the model for which the participant was unseen. Qualitatively, these maps corroborated known neuropathology in dementia, while still allowing for inter-individual variation (Supplementary Figure 2). We confirmed this apparent corroboration quantitatively by comparing a voxel-wise average map $\overline{R}_{dementia}$ (Supplementary Figure 3), containing positive relevance from all correctly predicted dementia patients, with a statistical reference map G (Supplementary Figure 4) from an activation likelihood estimation (ALE) meta-analysis [42]. For sanity checks, we also computed average maps from three alternative pipelines, $\overline{R}_{sex}$, $\overline{R}_{randomized\ weights}$and $\overline{R}_{randomized\ images}$. The comparisons with the reference map were done by binarizing the maps on both sides of the comparison at various thresholds and measuring the Dice overlap (Figure 2a). For the three alternative pipelines the amount of overlap decreased monotonically as the binarization threshold rose (Figure 2b), whereas for $\overline{R}_{dementia}$ it stabilized as the maps grew sparser, indicating more similarity with G. This effect was reaffirmed by a normalized cross-correlation [43] of 0.64 for $\overline{R}_{dementia}$, compared to 0.41, 0.40 and 0.12 of $\overline{R}_{sex}$,

$\overline{R}_{randomized\ weights}$ and $\overline{R}_{randomized\ images}$ respectively. In addition, we performed a region-wise, qualitative comparison of $\overline{R}_{dementia}$ and G, also yielding general agreement (Figure 2c), with the most important regions in both maps being the nucleus accumbens, the amygdala, and the parahippocampal gyrus. Next, we tested the importance of the detected regions in a predictive context, by applying an iterative mask-and-predict procedure. For each participant, we produced a baseline dementia-prediction $\hat{y}_0$ and relevance map R_{task} for each pipeline LRP_{task}. We then iteratively masked out the most important regions of the image according to the relevance map and recorded how the prediction changed as a function of the occlusion (Figure 2d). Using only true positives, the predictions should ideally start out at approximately 1.0 (empirically found to be 0.89 on average) and trend towards 0.5 (random prediction) as a larger proportion of the image is occluded. The rate of decline is indicative of whether the masked regions contain information essential for the classifier to classify the image correctly. Over 20 iterations we observed that the predictions based on maps from both $LRP_{dementia}$, LRP_{sex} and $LRP_{randomized\ weights}$ decreased, but $LRP_{dementia}$ at a distinctly steeper rate than the rest (Figure 2d). To quantify this observation we calculated an area over the perturbation curve (AOPC) of 0.231, 0.009, -0.001 and 0.002 for $LRP_{dementia}$, LRP_{sex}, $LRP_{randomized\ images}$, $LRP_{randomized\ weights}$ respectively. Taken together, these results demonstrate that our pipeline generates maps with relevance in brain regions associated with changes in dementia.

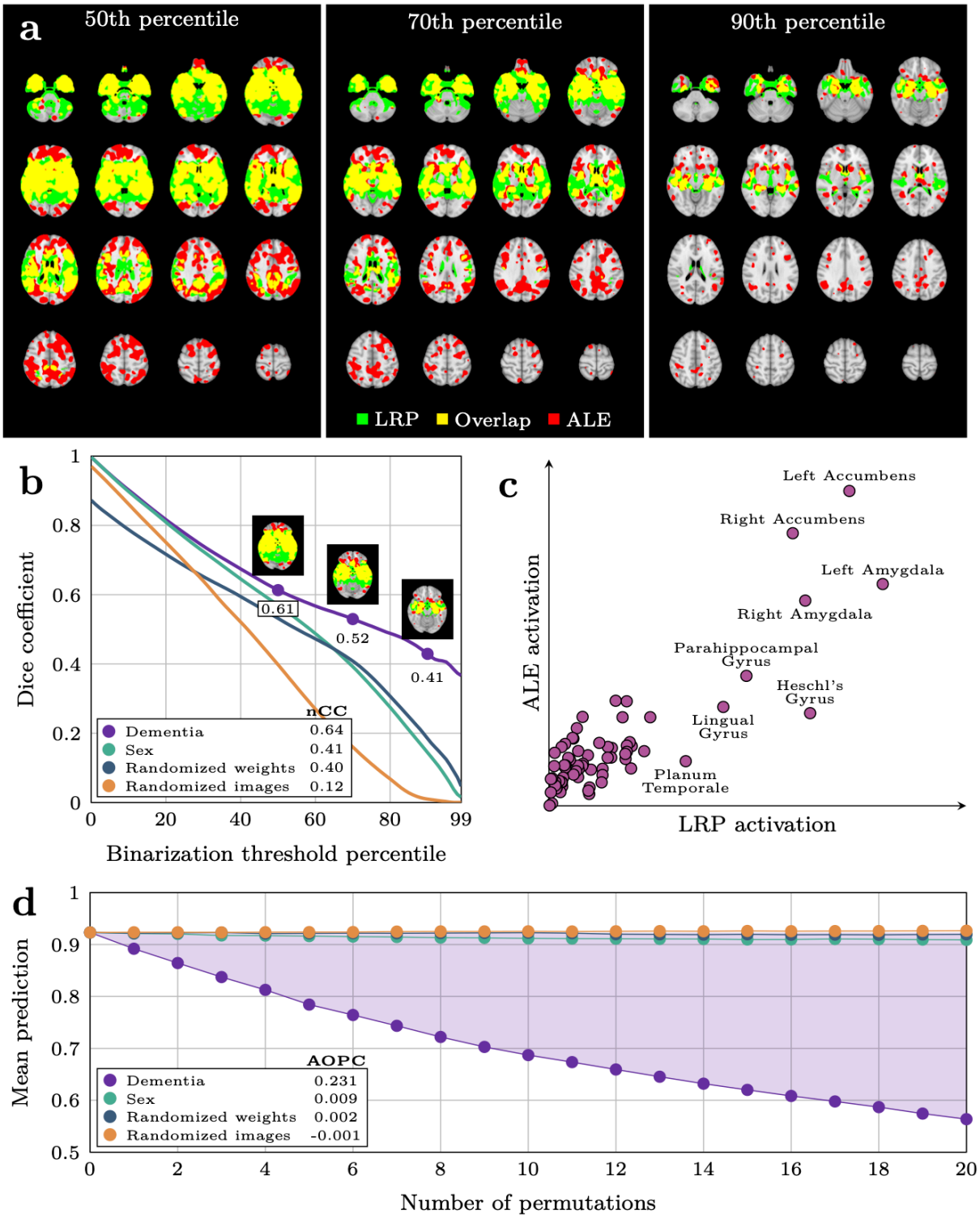


Figure 2: Validation of relevance maps from the dementia pipeline compared with three alternative pipelines. a *Visualization of the comparison between the binarized average relevance*

map $\overline{\boldsymbol{R}}_{\boldsymbol{dementia}}$ *from the dementia-pipeline and the binarized statistical reference map* $\boldsymbol{G}$ *from GingerALE, at different thresholds for binarization.* ***b*** *Overlap between the four average relevance maps* $\overline{\boldsymbol{R}}$ *from our four pipelines and* $\boldsymbol{G}$ *as a function of the binarization threshold. The numbers in the legend denote the normalized Cross Correlation (nCC) for each pipeline* ***c*** *Mean voxel-wise activation in* $\overline{\boldsymbol{R}}_{\boldsymbol{dementia}}$ *and* $\boldsymbol{G}$*, grouped by brain region.* ***d*** *Average participant-wise prediction from the dementia model after iteratively masking out regions of the image according to relevance maps from the four pipelines. Area over the permutation curve (AOPC) for the dementia map is indicated by the shaded area and denoted in the legend for all pipelines.*

Output from the explainable dementia pipeline has prognostic value for MCI patients

For the MCI patients (n=1256, timepoints=6448), previously unseen by all models, we built an averaging ensemble to procure a singular out-of-sample prediction and relevance map per patient per timepoint. Put together, we let this represent a morphological record (illustrated in Figure 4 and Supplementary Figure 5 and 6) visualizing the absolute quantity (indicated by the prediction) and location (indicated by the relevance map) of dementia-related pathology detected by the models over time. Qualitatively, both predictions and maps were relatively stable within a participant over time, while allowing enough variation to compose what resembled a trajectory. To investigate the prognostic value of our proposed morphological records we divided the MCI patients into three subgroups based on their trajectories in the follow-up period: those who saw improvement of their condition (n=80), those with a stable diagnosis throughout (sMCI, n=754), and those who progressed into dementia (pMCI, n=304). The remaining (n=118) had either a non-MCI diagnosis at the first timepoint, or a more complex diagnostic trajectory (e.g MCI -> AD -> CN), and were excluded from the subsequent analyses. We observed that the predictions in the first group were generally very low (mean $\hat{y}$ = 0.13, Supplementary Figure 7a), indicating that the models detected little, if any, evidence of dementia in these participants. For the stable patients the mean prediction was higher (mean $\hat{y} = 0.33$), but still below the classification threshold of 0.5, whereas in the progressive group the model predicted the average patient to already have dementia (mean $\hat{y} = 0.72$). Importantly, this was also true when considering only timepoints before these patients received the clinical diagnosis (mean $\hat{y} = 0.65$, Supplementary Figure 7b), suggesting that the model found evidence of the disorder before the clinical

symptoms surpassed the diagnostic threshold. To formally delineate the differences in predictions leading up to the potential diagnosis, we combined the improving and stable patients into a non-progressive group (nMCI, n=834), and sampled patients to match the progressive group based on their visiting histories, leading up to a terminal diagnosis timepoint (or a constructed non-diagnosis timepoint in the non-progressive group). In this matched dataset (n=550) we applied a linear mixed model controlling for age and sex and observed that the group difference was even greater than previously observed ($\beta = 0.47$, $p = 6.05 \times 10^{-71}$, Figure 3a, Supplementary Table 6). Furthermore, we observed a significant difference in longitudinal slopes ($\beta = 0.05$ increase in prediction per year, $p = 8.14 \times 10^{-17}$) indicating a greater rate of brain change detected by the model in those who would be diagnosed with dementia at a later point in time.

The large group differences in the dementia predictions leading up to a potential diagnosis suggests this as a biomarker with innate prognostic value, yet the most salient part of our morphological records were the relevance maps. Thus, we performed exploratory analyses based on these to further differentiate the non-progressive and progressive groups and characterize both inter- and intra-group heterogeneity. However, given the high dimensionality of the maps and the relatively small number of patients, we first applied a principal component analysis (PCA) to relevance maps from all MCI patients, effectively compressing their information content into a smaller set of characteristic variables encoding facets of the maps, enabling the subsequent analyses. We retained the 64 components that explained the largest amount of variance (Supplementary Figure 8) and observed that they qualitatively clustered into three overarching categories. The first component was a generic component detecting general presence of relevance, resembling the average map from dementia patients, and thus made up a cluster by itself. The next cluster was comprised of the subsequent three components that captured high level, abstract patterns of relevance, namely differences in lateralization, along the sagittal axis and in subcortical regions (Figure 3b). The final cluster consisted of the remaining 60 components that captured specific, intricate patterns of presence/non-presence of relevance in regions revealed in the previous analyses (Supplementary Figure 9). To investigate the potential of using the relevance maps for prognosis, we first performed a survival analysis using a Cox proportional hazards model where getting a diagnosis was considered the terminal event.

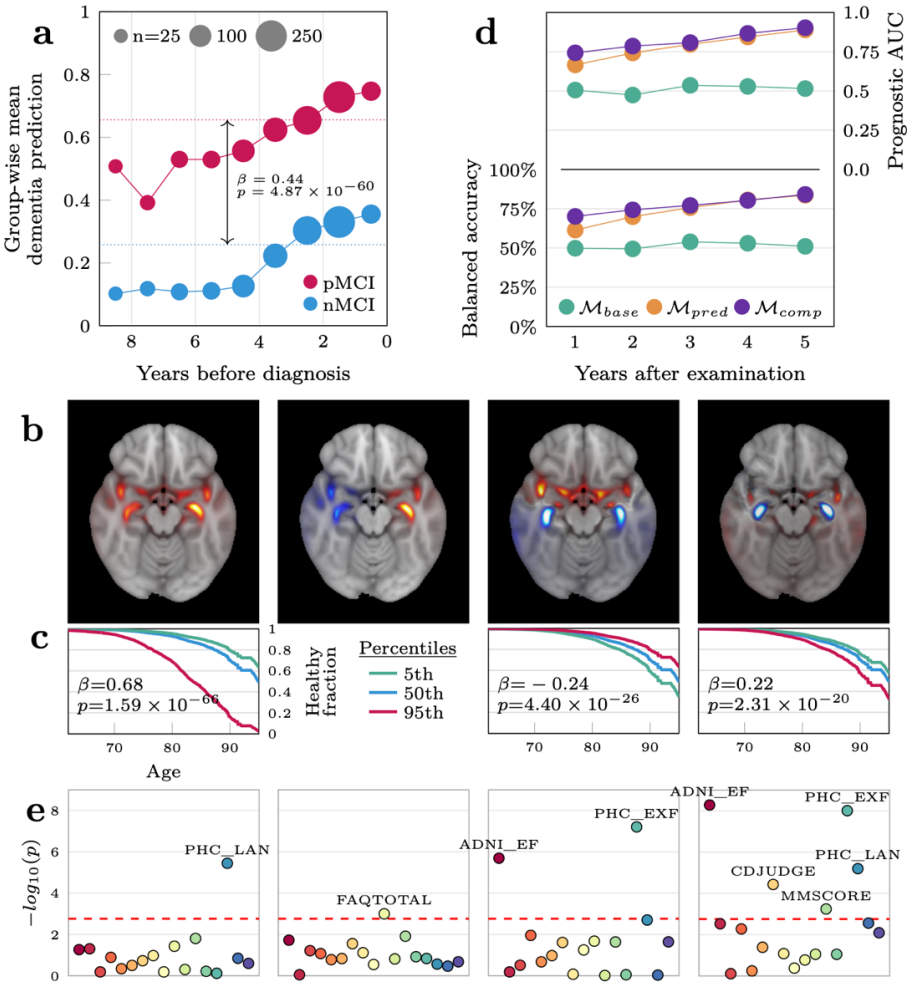


Figure 3: Utility of the dementia pipeline for predicting progression and characterizing individual-level deviations in the mild cognitive impairment cohort. ***a*** *Group-wise mean predictions from the dementia-model in the progressive and non-progressive groups in the years before a diagnosis was given.* ***b*** *The four first voxel-wise components of the principal component analysis plotted in MNI152-space.* ***c*** *Survival curves for the average MCI patient (blue) and fictitious patients at the extreme percentiles of the span for each component. The second component was not significant and is not shown.* ***d*** *Predictive performance of the three models predicting progression in the years following the MRI examination. The baseline model ($\mathcal{M}_{base}$) included only sex and age as covariates, the next model $\mathcal{M}_{pred}$ included the prediction from the dementia classifier as a predictor, while the final model $\mathcal{M}_{comp}$ also added the component vectors representing the relevance maps.* ***e*** *Significance levels of correlations between the each of the four PCA components and various cognitive measures. The six annotated measures are*

composite language (PHC_LAN) and executive function (PHC_EXF) scores from the ADSP Phenotype Harmonization Consortium, total score from the Functional Activities Questionnaire (FAQTOTAL), composite executive function score from UW – Neuropsych Summary Scores (ADNI_EF), clinical evaluation of impairment related to judgement and problem solving (CDJUDGE) from the Clinical Dementia Rating, and an overall measure of cognition from the Mini-Mental State Examination (MMSCORE).

Specifically, we modelled the fraction of the population without a diagnosis as a function of age and used the subject-wise loadings of c_t as predictors. After Benjamini-Hochberg correction, 37 of these components were significantly associated with staying undiagnosed (Figure 3c and Supplementary Table 7). However, we observed a correlation between the singular dementia prediction $\hat{y}$ and the absolute magnitudes of these components (Supplementary Figure 10), indicating that the associations in the survival analysis could be induced by differences in the prediction rather than variability in the relevance maps. To mitigate this concern, we fit an equivalent model while stratifying on $\hat{y}$, observing that 29 associations remained significant, and that all coefficients had the same sign. Nonetheless, this analysis did not account for the predictions and relevance maps changing within a participant over time, so we reframed the question in a purely predictive setting, constructed to bear resemblance to a clinical scenario, using the same participants (nMCI=834, pMCI=304, total n=1138). For each MCI patient p at each timepoint t we asked whether we were able to predict, at yearly intervals γ up to five years into the future, whether p had progressed into dementia, using information from $LRP_{dementia}$ available at t. Importantly, all timepoints for all these participants were unseen by the dementia-model, yielding out of sample predictions and relevance maps from $LRP_{dementia}$, and we employed nested cross-validation to ensure the progression predictions were also out-of-sample. First, we fit a baseline model $\mathcal{M}_{base}$ with age and sex as predictors, showing no predictive efficacy at any timepoint (all AUCs ≈ 0.5, Supplementary Table 8), indicating that the dataset was not biased with respect to these variables. When adding the prediction from the dementia model $\hat{y}_t$ as a predictor in model $\mathcal{M}_{pred}$ we saw large improvements in prognostic efficacy at all yearly intervals, culminating with a fold-wise mean AUC of 0.889 after five years (Figure 3d). In the final model, $\mathcal{M}_{comp}$, also including the component vector c_t as predictors, we saw further improvements for all years, peaking at 0.903 after five years ($p = 0.035$ when compared to

$\mathcal{M}_{pred}$ in a Wilcoxon signed-rank test across the outer folds). Overall, our best performing model predicted progression to dementia after five years with an AUC of 0.903, an accuracy of 84.1%, a positive predicted value of 0.92, a sensitivity of 0.82 and a specificity of 0.86 (Table 2).

Model	AUC	Balanced accuracy	PPV	Sensitivity	Specificity
$\mathcal{M}_{base}$	0.515	51.05%	0.14	0.09	0.93
$\mathcal{M}_{pred}$	0.889	83.61%	0.91	0.83	0.84
$\mathcal{M}_{comp}$	0.903	84.1%	0.92	0.82	0.86

Table 2: Predictive performance of the three models predicting progression five years into the future. *The baseline model $\mathcal{M}_{base}$ used only age and sex as covariates. $\mathcal{M}_{pred}$ also added the prediction from the dementia model at the current timepoint as a predictor, while $\mathcal{M}_{comp}$ additionally included the component vector c_t encoding information from the relevance maps.*

Facets of the relevance maps are associated with cognitive impairments in distinct domains

Finally, we tested whether common features found in the relevance maps, represented by the PCA component, were correlated with impairments in distinct cognitive and functional domains. We extracted 17 summary measures from 7 neuropsychological tests (Supplementary Table 9 and 10), performed approximately at the same time as an MRI examination, and tested for associations with the subject-wise loadings of c_t in 733 MCI patients using linear models. After FDR correction, while correcting for age, sex and $\hat{y}$, we found 48 significant correlations between 18 unique components and 14 of the cognitive measures (Figure 3e). Component 30 and the aggregate score from the Functional Activities Questionnaire (FAQTOTAL) had the highest number of significant hits among the components and measures respectively, both with six passing the threshold. Most importantly, the components showed distinct patterns of associations with the different cognitive measures (Supplementary Table 13). To ensure the significant associations were not driven by collinearity between components c_i and $\hat{y}$, we ran an equivalent analysis without including $\hat{y}$ as a predictor, observing that only 5/48 of the previously significant hits had coefficients with the opposite sign (Supplementary Table 14). To summarize, the spatial features captured in our relevance maps, and subsequently in our component vectors, were

associated with distinct patterns of performance on neuropsychological tests relevant for characterizing phenotypic heterogeneity in dementia patients (Supplementary Figure 11).

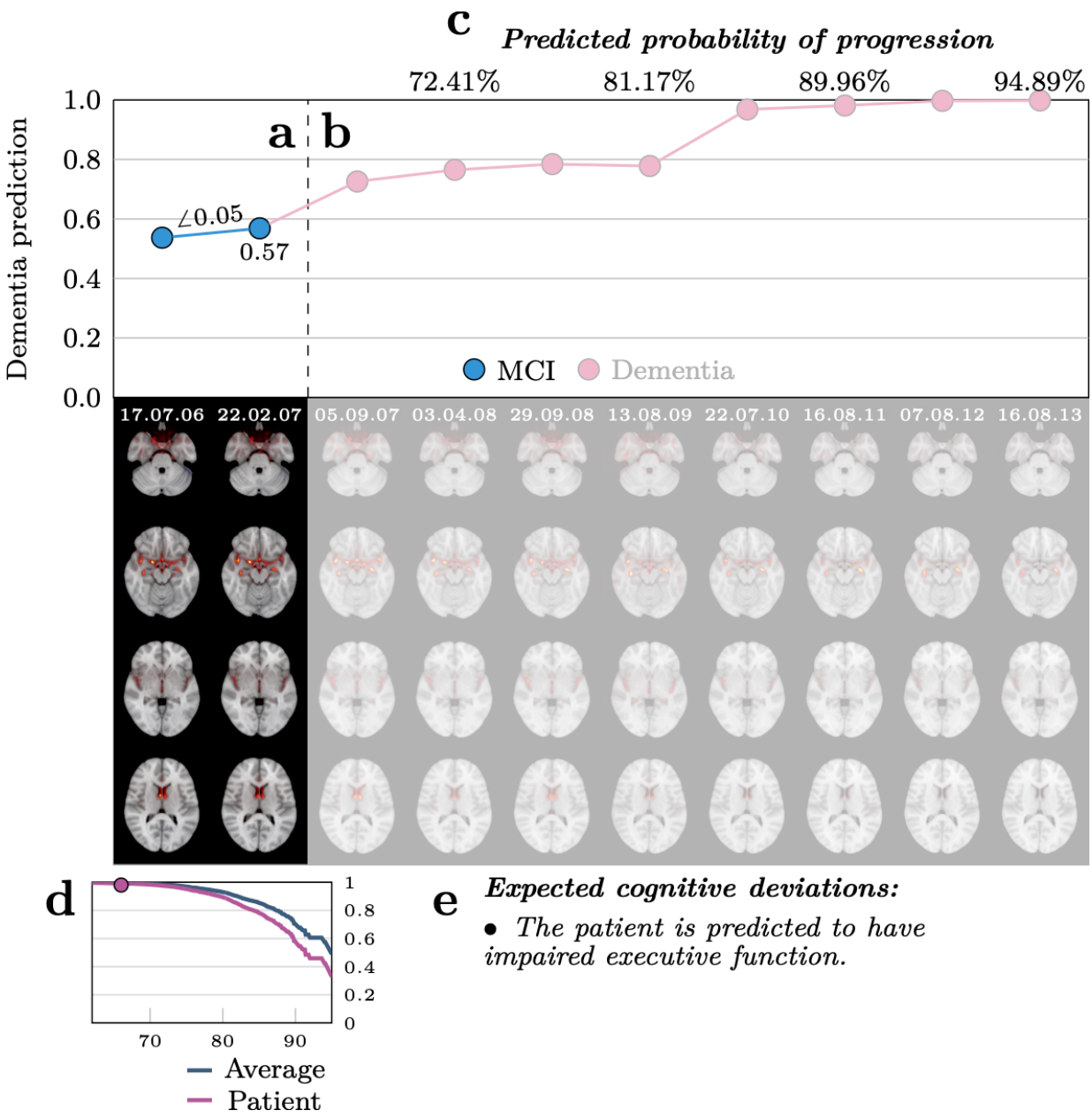


Figure 4: A visualization of the proposed morphological record for a randomly selected progressive MCI patient that was held out of all models and analyses. a *The top half shows the prediction from the dementia model at each visit, while the bottom part displays the relevance map underlying the prediction. The opaque sections (including* ***c****,* ***d,*** *and* ***e****) contain information accessible at the imagined current timepoint (22.02.07) to support a clinician in a diagnostic procedure. The angle (∠) represents the change in dementia prediction per year based on the first two visits.* ***b*** *Translucent regions reveal the morphological record for the remaining follow ups in the dataset, thus depicting the future. The ground truth diagnostic trajectory is encoded by*

the colour of the markers. ***c*** *Predicted probabilities of progression at future follow-ups based on the prediction and relevance map at the current timepoint.* ***d*** *Survival curve of the patient compared to the average MCI patient calculated from the prediction and relevance map. The marker indicates the location of the patient at the current timepoint.* ***e*** *A list of cognitive domains where the patient is predicted to significantly differ from the average based on the prediction and relevance map.*

Discussion

Given the huge burden of disease and expected increase in prevalence, innovative technological solutions for clinical decision making in dementia diagnostics and prognostics is urgently needed. Although commonly referred to as a homogenous condition or split into a few subtypes based on aetiology or pathophysiology [17], dementia patients exhibit unique and complex deficiencies, disease trajectories, and cognitive deficits. To explore the potential of brain MRI and XAI to characterize heterogeneity in the brain underpinnings of dementia, we trained neural networks to differentiate dementia patients from healthy individuals, and derived relevance maps using Layerwise Relevance Propagation to explain the individual-level decisions of the classifier. The relevance maps were specific to the individual, spanned regions that were predictive of dementia and corroborated existing knowledge of the anatomical distribution of neuropathology. In a cohort of MCI patients, it enabled characterization and differentiation of individual-level disease manifestations and trajectories linked to cognitive performance in multiple domains. While further validations in clinical contexts are needed, our XAI pipeline for dementia has the potential to be employed by clinicians for monitoring and characterizing disease development at the level of the individual subject.

There is a multitude of XAI techniques available for explaining the decisions of an image classifier, many of which have yielded promising results for dementia classification [38]. We employed LRP due to its straightforward interpretation as well as earlier studies indicating robustness [44], both properties we would consider integral in a clinical decision support system. But while procuring explanations that are *ipso facto* meaningful is an important step towards adoption of AI in clinical neuroimaging, it is not in itself sufficient. There is a host of predictive models that are trivially explainable, but not understandable [45], and there is genuine concern that XAI will lead to another level of systems that are formally well-defined, but opaque and obscure, and thus practically useless [46]. Thus, empirical explorations are imperative to investigate the nature of these explanations, examine how they may be useful and build essential trust [47]. In our validation, we observed that the explanatory maps produced by the dementia pipeline were more predictive and showed distinctly more agreement with existing knowledge of pathology than those produced by the three alternative pipelines. Given limitations that have been exposed in such methods earlier [48,49] these validations are crucial, and observing that our results both

corroborate earlier evidence [40] and extend beyond that, provides confidence that the explanations derived from the model are meaningful. However, we emphasize that the ultimate validation should happen in actual implementations of the technology in end-user systems, with clinical personnel applying it in clinical scenarios on realistic data.

We extend upon validating the relevance maps by proposing them as a potential epistemic and clinical tool to characterize individual facets of dementia. To this end, we explored if the maps contributed to predicting imminent progression from MCI to dementia, and correlated them with different cognitive measures, extending upon the current literature [38]. In both analyses we found evidence, although modest, that the maps are informative beyond the predictions of the model. To illustrate the potential of the pipeline for clinical decision making we compiled its output into a proposed morphological record (visualized for a single patient in Figure 4) that can help clinicians localize morphological abnormalities during a diagnostic process. Identifying subtle pathophysiology through deep phenotyping could have a huge potential for charting the heterogeneity of dementia, providing precise biological targets to guide future research. Furthermore, for the individual patient, it can support personalized diagnosis to identify appropriate disease-modifying treatments, and in the future, hopefully, accurate therapeutic interventions.

The regions with the highest density of relevance in our maps were the nucleus accumbens, amygdala and the parahippocampal gyrus, all of which are strongly affected in dementia [50–52]. While the two latter corroborate the established involvement of the medial temporal lobe [53] it is surprising that the hippocampus does not appear in our analyses, as it has frequently in similar studies [38]. While this could be caused by actual localization of pathology [54] we consider it more likely to be related to the inner working of the model. Specifically, the CNN relies on spatial context to identify brain regions before assessing their integrity, utilizing filters that span areas of the image larger than those containing the region itself. In the backwards pass, LRP uses these filters, and thus the localization of relevance is not necessarily voxel precise. Furthermore, we believe the model broadly can be seen as an atrophy detector, which necessarily entails looking for gaps surrounding regions instead of directly at the regions themselves. Therefore, while the

relevance maps provide important information, they depend on contextual information and thus rely on interpretation from clinicians to maximize their utility in clinical practice.

We focused our analyses mainly on the relevance maps, but the results with largest, immediate, potential for clinical utility were the predictions from the dementia classifier. Other studies have shown the efficacy of machine learning models in differentiating dementia patients and healthy controls [28], but it is intriguing that we see a large discrepancy in the predictions of the progressive and non-progressive MCI patients many years before the dementia diagnosis is given. This corroborates findings from theory-driven studies [55] and a recent deep learning study [27], implying detectable structural brain changes many years before the clinical diagnosis is given. This gives hope for advanced technology to contribute to early detection and diagnosis through MRI based risk scores, in our case supported by a visual explanation. If curative treatments prove efficacious and become accessible, early identification of eligible patients could be imperative [56]. Furthermore, timely access to interventions have shown efficiency in slowing the progress of cognitive decline [57], in addition to improving the quality of life for those afflicted and their caregivers [26,58]. Widely accessible technology that allows for early detection with high precision could play a key role in the collective response to the impending surge of patients and provide an early window of opportunity for more effective treatments.

While our results show a great potential for explainable AI, and particularly LRP, as a translational technology to detect and characterize dementia, there are limitations to our study. First, there are technical caveats to be aware of. Most importantly, there is an absolute dependence between the predictions of our model and the relevance maps. In our case, when we qualitatively assessed the relevance maps of the false negatives, they were indistinguishable from the true negatives. This emphasizes the fact that when the model is wrong, this is not evident from the explanations. Next, while the maps contain information sufficient to explain the prediction, they are not necessarily complete. Thus, they don't contain all evidence in the MRI pointing towards a diagnosis, a property which could prove essential for personalization. We have addressed this problem through pragmatic solutions, namely ensembling and targeted augmentations, but theoretical development of the core methodology might be necessary to theoretically guarantee complete maps. Beyond the fundamental aspects of LRP, there are

weaknesses to the present study that should be acknowledged. First, the dataset with dementia patients portrayed as heterogeneous mostly consists of ADNI and OASIS data, and thus patients with a probable AD diagnosis (although clinically determined). Thus, while we consider it likely, it is not necessarily true that the dimension of variability spanning from healthy controls to dementia patients portrayed by our model has the expressive power to extrapolate to other aetiologies. To overcome this in actual clinical implementations, we encourage the use of datasets that are organically collected from subsets of the population that are experiencing early cognitive impairments, for instance from memory clinics. A related problem is the out-of-sample generalization, especially related to scanners and acquisition protocols. Although we utilize data from many sites, which we have earlier shown to somewhat address this problem [59], in combination with transfer learning, we did not explicitly test this by e.g., leaving sites out for validation. Again, we advise that clinical implementations should be based on realistic data, and thus at least be finetuned towards data coming from the relevant site, scanner, and protocol implemented in the clinic [60]. This also includes training models with class frequencies matching those observed in clinical settings, instead of naively balancing classes as we have done here. Next, we want to explicitly mention the cyclicality of our mask-and-predict validation. In a sense it trivially follows that regions that are considered important by a model are also the ones that are driving the predictions, and thus it is no surprise that the relevance maps coming from the dementia model are more important to the dementia model than the maps coming from e.g., the sex model. We addressed this by alternating the models for test and validation, but fully avoiding this circularity would require disjunct datasets, and more and larger cohorts. In summary, the predictive value for the individual patient must be interpreted with caution. However, our extensive validation approach as well as our thorough explanation of the method and its limitations, and training on a massive dataset, provides a first step towards making explainable AI relevant for clinical decision support in neurological disorders. Nonetheless, it also reveals a complicated balance between validating against existing knowledge and allowing for new discoveries. In our case, confirming whether small details revealed in the relevance maps are important aspects of individualization or simply intra-individual noise requires datasets with a label-resolution beyond what currently exists. Thus, we reiterate our belief that the continuation of our work should happen at the intersection between clinical practice and research [61], by

continuously collecting and labelling data to develop and validate technology in a realistic settings.

To conclude, while there are still challenges to overcome, our study provides an empirical foundation and a roadmap for implementations of brain MRI based explainable AI in clinical decision support systems for personalization. Specifically, we show that deep neural networks trained on a heterogenous set of brain MRI scans can predict dementia, and that their predictions can be made human interpretable. Furthermore, our pipeline allows us to reason about neurobiological aberrations in individuals showing early signs of cognitive impairment by providing personalized characterizations which can subsequently be used for precise phenotyping and prognosis, thus fulfilling a realistic clinical purpose.

Materials and Methods

Data

All data used in the present study have been obtained from previously published studies which have been approved by their respective institutional review board or relevant research ethics committee.

To train the dementia models we compiled a case-control dataset from seven different sources (Supplementary Table 1), consisting of patients with a dementia diagnosis and healthy controls from the same scanning sites. Because of the different diagnostic criteria used in the original datasets we applied different rules to achieve a singular, heterogeneous dementia label (Supplementary Table 2). We extracted all participants with a dementia-diagnosis at all timepoints to comprise the patient group (n=854). Then, for each unique proxy site (In ADNI, due to the large number of scanners and acquisition protocols, and the work put into unifying them, we used field strength as a proxy for site), sex, and age-bin spanning 10 years, we sampled an equal number of healthy controls to form the matched control set (total n=1708, Table 1). Lastly, before modelling, we split the data into five equally sized folds stratified on diagnosis, site, sex, and age, such that all timepoints for a single participant resided in the same fold.

For the MCI dataset we started with all participants from all ADNI waves with an MCI diagnosis (subjective memory complaint, MMSE between 24 and 30, CDR>0.5 with memory box>0.5, Weschler Memory Scale-Revised <9 for 16 years of education, <5 for 8-15 years of education and <3 for 0-7 years of education) [62], on at least one timepoint. These were 12661 images from 6448 visits for 1256 participants, none of which were used for model training. This selection criterion ensured all participants had an MCI diagnosis at one point in time, though it did not limit us to only those timepoints. Thus, in addition to those with a consistent, stable, MCI diagnosis (sMCI), we had a variety of diagnostic trajectories, including those transitioning from normal cognition to MCI, MCI to AD (pMCI) and various other combinations. Before the subsequent analyses we discarded all participants without an MCI diagnosis initially, and everyone with ambiguous trajectories (e.g. MCI->CN->AD), leaving 5607 visits from 1138 participants.

Modelling

All dementia models were variants of the PAC2019-winning simple fully convolutional network (SFCN) architecture [63,64], modified to have a single output neuron with a sigmoid activation. The architecture is a simple, VGG-like convolutional neural network with 6 convolutional blocks and approximately 3 million parameters. We initialized the model with weights from a publicly accessible brain age model previously shown to have superior generalization capabilities when dealing with unseen scanning sites and protocols [59]. Before modelling, all images were skullstripped [65] and linearly registered to MNI152 space [66] using a previously developed pipeline [59] relying on FreeSurfer v5.3 and FSL v6.0 [67]. The models were trained on a single Nvidia A100 GPU with 40GB of memory, Tensorflow 2.6 [68] through the Keras interface [69]. We used a vanilla stochastic gradient descent (SGD) optimizer with a learning rate defined by the hyperparameter settings (see next section), optimizing the binary cross-entropy loss. All models ran for 160 epochs with a batch size of 6, and for each run the epoch with the lowest validation loss was chosen. Varying slightly depending on the hyperparameters, a single model trained in approximately 4 hours.

For each hold-out test fold we trained models on three of the remaining folds and validated on the fourth, akin to a cross-validation with an additional out-of-sample test set, to achieve out-of-sample predictions for all 1708 participants while allowing for hyperparameter tuning. The

hyperparameters we optimized were dropout $d \in \{0.25, 0.5\}$ and weight decay $w \in \{10^{-2}, 10^{-3}\}$. Additionally, we tested stepwise, one-cycle and multi-cycle learning rate schedules (Supplementary Figure 12b), and a light and a heavy augmenter (Supplementary Table 12). Initial values for the learning rate were set manually based on a learning rate sweep [70], though kept conservative to preserve the learned features from the pretraining (Supplementary Figure 12a). The hyperparameter search was implemented as a naive grid-search over the total 24 different configurations (Supplementary Figure 13). We selected the model procuring the best AUC in the validation set to produce out-of-sample predictions for the outer hold-out fold. In the final evaluation of the models, we compiled predictions for all participants, for each using the model where they belonged to the hold-out test set. Our main method for measuring performance was the AUC, but we also report accuracy, which, due to our matching procedure, is equivalent to balanced accuracy.

Relevance maps

We built a pipeline $LRP_{dementia}$ for generating relevance maps by implementing LRP (Bach et al., 2015) on top of the trained classifier. LRP is a technique for explaining single decisions made by the model, and thus, when running the pipeline on input X a relevance map R is generated alongside the prediction $\hat{y}$. R is a three-dimensional volume, representing a visual explanation for $\hat{y}$, where each voxel $r_{i,j,k} \in R$ has a spatial position i, j, k corresponding to the location of an input voxel $x_{i,j,k} \in X$. Furthermore, the intensity of $r_{i,j,k}$ can be directly interpreted as how much voxel $x_{i,j,k}$ contributes to $\hat{y}$, such that $\sum_{r \in R} r = \hat{y}$. In the original LRP-formulation, relevance r is propagated subsequent layers Z_0 and Z_1 according to the relative contribution of one artificial neuron $a_i \in Z_0$ in the first layer on relevance in all artificial neurons $a_j \in Z_1$ in the following layer,

$$r(a_i) = \sum_j \frac{a_i w_{ij}}{\sum_k a_k w_{kj}} r(a_j),$$

where w_{ij} denotes the weight between a_i and a_j. We controlled the influence of different aspects of the explanations using a composite LRP strategy [41], combining different formulations of the LRP-formula for the different layers in the model to enhance specific aspects of the relevance

maps. Specifically, we employed LRP_{ϵ} for the prediction layer to retain the most salient explanations,

$$r_{\epsilon}(a_i) = \sum_j \frac{a_i w_{ij}}{\epsilon + \sum_k a_k w_{kj}} r(a_j).$$

For the central convolutional layers, we upweighted positive relevance (e.g. features increasing the prediction, corresponding to evidence for a diagnosis) with $\text{LRP}_{\alpha\beta}$,

$$r_{\alpha\beta}(a_i) = \sum_j \left(\alpha \frac{(a_i w_{ij})^+}{\sum_k (a_k w_{kj})^+} - \beta \frac{(a_i w_{ij})^-}{\sum_k (a_k w_{kj})^-} \right) r(a_j),$$

Where $(\cdot)^+$ and $(\cdot)^-$ denote positive and negative contributions respectively. For the input layer and the following convolutional layer we employed LRP_{b} (also denoted as LRP_{flat}), to smooth finer details of the relevance maps,

$$r_{\text{b}}(a_i) = \sum_j \frac{1}{\sum_k 1} r(a_j).$$

The configuration of the full strategy can be found in Supplementary Table 5. The raw relevance maps produced by the pipeline were full brain volumes with the same dimensionality as the MRI data (167x212x160 voxels) containing mostly (see below) positive relevance.

Notation-wise we generally consider the relevance map $R(X_i)$ for an image X_i to be a function of the model m_{task}, where $task$ indicates which task the model was trained for, the LRP strategy $\text{LRP}_{\text{composite}}$ and the image X_i

$$R(X_i) = f\left(m_{task}, \text{LRP}_{\text{composite}}, X_i\right).$$

Because the composite LRP strategy described above is kept fixed in our pipeline, we contract this to

$$R(X_i) = f(m_{task}, X_i).$$

Furthermore, we let the model-specifier task annotate the map for a further simplification

$$R_{task}(X_i) = f(X_i).$$

Thus, LRP_{task} is used to annotate the full pipeline for a given task, while $R_{task}(X_i)$ denotes a single relevance map generated by this pipeline for image X_i. When the task is given by the context, we sometimes simplify this further to $R(X_i)$, and when a general image is considered, we simply use R to denote its relevance map.

While we generally discuss our pipeline as a singular one, there were in reality five approximately equivalent pipelines (corresponding to the models trained for the five test folds), and which one is used depends on what image was used as input. Specifically, for each participant diagnosed with dementia, the pipeline is chosen where the participant was part of the hold-out test set while training the model, and both the relevance maps and the predictions are thus always out-of-sample. For participants used in the MCI analysis, which are all out-of-sample for all models, we created an ensemble by averaging the predictions and the voxel-wise relevance across all models.

Before implementing the LRP procedure we made two slight modifications to the models to facilitate the backwards relevance propagation, both leaving the functional interface of the model unchanged. First, we removed the sigmoid activation in the final layer, so that the output of the model changed from a bounded continuous variable $\hat{y} \in [0, 1]$ to an unbounded prediction $\hat{y}_\sigma \in [-\infty, \infty]$. In this space a raw prediction of $\hat{y}_\sigma = 0$ is equivalent to a sigmoid-transformed prediction of $\hat{y} = 0.5$, and thus $\hat{y}_\sigma < 0$ means that the model predicts control status for the given participant, and oppositely $\hat{y}_\sigma > 0$ implies that the model predicts a dementia diagnosis. Furthermore, this means that all positive relevance $r \in R, r > 0$ can be interpreted as visual evidence in favour of a dementia diagnosis. Secondly, we modified the model by fusing all batch normalization layers with their preceding convolutional layers, adjusting their weights and biases to match the shift and scaling previously performed by the normalization layer [71,72].

After generation, the relevance maps are in the same stereotaxic space as their corresponding, linearly registered, input MRIs. To ensure intra-individual comparisons were done in the same space we non-linearly registered the maps to MNI152- space before subsequent statistical analyses were run. First, we registered the preprocessed MRIs X_i used as inputs to the 1mm MNI152 template packaged with FSL using fnirt with splineorder=2. We then applied the transformation computed for X_i to $R(X_i)$ using applywarp. We also restrained our relevance maps to contain strictly positive relevance, evidence in favour of a dementia prediction, by clipping them to a minimum value of 0. Furthermore, to remove edge-effects from our analyses, we enforce that there is no relevance in non-brain tissue by nullifying all relevance outside the brain:

$$\forall(i,j,k)\left[x_{i,j,k} = 0 \;\Rightarrow\; r_{i,j,k} = 0\right].$$

All visualized relevance maps are plotted after non-linear registration, overlayed on the MNI152-template. As the maps are three-dimensional, we generally plot a collection of distributed axial slices. The relevance is coloured by the nibabel v3.2.2 [73] cold_hot colourmap. Since the absolute relevance values vary between maps, all maps are normalized to the intensity range [0, 1] in the visualizations.

Validating the relevance maps

Earlier studies have shown that interpretability techniques in general are prone to generate visual explanations that do not capture salient parts of the input [48,49]. To investigate the extent of this for our pipeline $LRP_{dementia}$ we employed two analyses to assess the sanity of the relevance maps. The first was a domain-specific analysis comparing the relevance maps to existing knowledge of the pathology of dementia, while the second was a purely quantitative analysis examining how important the regions found by the pipeline are for the dementia prediction $\hat{y}$. In both cases we contrasted the relevance maps generated from the main pipeline with three alternative pipelines representing variants of a null hypothesis, all expected to produce relevance maps with no significant association to dementia.

$LRP_{random\ images}$ represents the simplest alternative pipeline, and is built around the dementia-model, but with an additional preprocessing step scrambling the input,

$$R_{random\ images}(X_i) = R_{dementia}(\mathcal{X}_i), \text{ where}$$
$$\mathcal{X}_i = \mathcal{N}(\overline{X}_i, \sigma_{X_i}).$$

$LRP_{random\ images}$ is expected to generated relevance maps where the relevance is evenly distributed across the entire image. In the next pipeline $LRP_{random\ weights}$ we replaced the dementia-model with a model with random weights,

$$R_{random\ weights}(X_i) = R(m_\theta, X_i).$$

m_θ has not been trained for any task, and thus has random weights initialized by the default Keras "*Glorot Uniform*" weight-initializer. This pipeline is expected to produce relevance maps which correlate with the raw voxel intensities, e.g. high intensity in the input should entail more (absolute) relevance, thereby reflecting aspects of morphology. The final and most realistic alternative pipeline was LRP_{sex}, where we replaced the dementia-model with a binary sex-classifier,

$$R_{sex}(X_i) = R(m_{sex}, X_i).$$

The sex-classifier was trained to differentiate males from females in one of the splits from the dementia-dataset, achieving an out-of-sample AUC of 0.956 and a balanced accuracy of 89.40%. We did not do any hyperparameter optimization for this model but used the best configuration from the dementia cross-validation in the same fold. The heatmaps from this pipeline should reflect regions where there is intra-individual variation in morphology, which are predictive of sex but with minimal association with dementia.

As a proxy for existing knowledge in the literature we performed an ALE meta-analysis using Sleuth v3.0.4 [74] and GingerALE v3.0.2 [42]. We used Sleuth to search for relevant articles with the query

Imaging Modality is MRI

AND

Context is disease

AND

Diagnosis is Dementia OR Alzheimer's Disease OR Lewy Body Dementia OR Frontotemporal Dementia OR Non-Aphasic Frontotemporal Dementia

in the Voxel-based morphometry database, yielding 394 experiments from 124 articles. These experiments contained 3972 foci, 280 of which were outside the MNI152 mask, leaving 3692 to be loaded into GingerALE. Then the reference map G, with voxels $g_{i,j,k}$, was generated by an ALE meta-analysis using the default parameters: Cluster-level FWE=0.01, Threshold Permutations=1000, P Value=0.001. The reference map is visualized in Supplementary Figure 4.

We performed four pairwise comparisons to estimate the amount of overlap between each of the pipelines and G. For each pipeline, we first computed an average relevance map $\overline{R}$ across all true positives (e.g. dementia patients where the dementia-model correctly predicted a diagnosis, n=697, Supplementary Figure 1c), by computing their voxel-wise average relevance. Next, we binarized both the average map and the reference map by thresholding them at multiple percentiles $p \in [0, 100)$,

$$\bar{R}_p = \begin{cases} 1 & r_{i,j,k} > percentile(r,p) \\ 0 & else \end{cases},$$
$$G_p = \begin{cases} 1 & g_{i,j,k} > percentile(g,p) \\ 0 & else \end{cases}.$$

Then, for each percentile p we calculate the Sørensen-Dice coefficient SDC between the two,

$$SDC_p(\bar{R}_p, G_p) = \frac{\sum_{i,j,k} r_{i,j,k} g_{i,j,k}}{\sum_{i,j,k} r_{i,j,k} + \sum_{i,j,k} g_{i,j,k}}, r \in \bar{R}_p, g \in G_p$$

which allows us to plot the Dice coefficient as a function of the percentile (Figure 2b). To have a singular numerical basis for comparison we also computed the normalized cross-correlation [43] between the (non-binarized) average maps $\overline{R}$ and the reference map G ,

$$nCC(\bar{R}, G) = \frac{\sum_{i,j,k}(r_{i,j,k} - \bar{r})(g_{i,j,k} - \bar{g})}{\sqrt{\sum_{i,j,k}(r_{i,j,k} - \bar{r})^2 * \sum_{i,j,k}(g_{i,j,k} - \bar{g})^2}}, r \in \bar{R}, g \in G.$$

To facilitate an intuitive understanding of what parts of the brain the dementia-model is focusing on, we also performed a similar, region-wise comparison. This was done by extracting a subset of voxels from the average relevance map $\overline{R}_{dementia}$,

$$\overline{R}_\rho = \{r_{i,j,k} \, \mathrm{I} \, (i, j, k) \in \rho\},$$

where ρ is one of 69 regions defined in the Harvard-Oxford cortical and subcortical atlases [75]. We did the same for G and let the mean activation per region for both constitute a tuple

$$\left(\frac{\sum_{r \in R_\rho} r}{|R_\rho|}, \frac{\sum_{g \in G_\rho} g}{|G_\rho|}\right)$$

plotted Figure 2c. However, since it is non-trivial to determine which aggregation method corresponds to the most understandable and intuitive interpretation, we also created plots for tuples of sums,

$$\left(\sum_{r \in R_\rho} r, \sum_{g \in G_\rho} g\right)$$

and maximum values

$$\left(\max_{r \in R_\rho} r, \max_{g \in G_\rho} g\right)$$

per region in Supplementary Figure 14.

To quantify the importance of the spatial locations captured by the various LRP pipelines for predicting dementia, we implemented a procedure for iteratively occluding parts of the image based on the relevance maps and observing how the prediction from the dementia model changed [76]. Still using the true positives, for each pipeline LRP_{task} for each MRI X_0 we generated a baseline dementia-prediction $\hat{y}_0$ and relevance map R_{task}. Then we located the voxel with the highest amount of relevance in R_{task} and replaced a 15x15x15 cube centred around the voxel with random uniform noise $\mathcal{U}(0,\ 1)$, effectively concealing all information contained in this region. Next, we ran the modified image X^1_{task} through the dementia-model to see how the prediction $\hat{y}^1_{task}$ changed as a function of the occlusion. Note that injecting a box of random noise into the image is not trivially equivalent to removing information, however we specifically applied the same modification in the random box-augmentation during training and are thus hopeful that the model is invariant to the injection beyond the information removal. We iteratively applied this modify-and-predict procedure, also masking out the regions from the relevant maps between each iteration to minimize overlap of occlusion windows, for 20 iterations, producing a list of predictions $[\hat{y}_0,\ \hat{y}^1_{task}, \hat{y}^2_{task}, \ldots, \hat{y}^{19}_{task}]$ plotted for each pipeline in Figure 2d (averaged across all true positives). The rate of decline in these traces indicate the importance of the regions found in the respective relevance maps. We quantified the differences between the pipelines LRP_{task} by calculating the area over the area over their perturbation curves [76],

$$AOPC_{task} = \frac{1}{20}\left(\sum_{i=1}^{20} \hat{y}_0 - \hat{y}^i_{task}\right).$$

Exploratory analyses in the MCI cohort

In the exploratory MCI analyses we used $LRP_{dementia}$ to generate predictions and relevance maps for participants from ADNI who were given an MCI diagnosis at inclusion. We first compiled the predictions and relevance maps (and the corresponding timestamps) for each participant at all timepoints into a single data structure we called a morphological record. We then tried to utilize this data structure to differentiate three groups: stable MCI patients (sMCI), progressive MCI patients (pMCI), and those who saw improvement in their cognition throughout the data collection phase. The remaining participants, e.g. those who either passed through all three diagnostic stages, or bounced between diagnoses, were excluded. Furthermore, we combined the stable and improving cohorts into a non-progressive group (nMCI) to facilitate binary group comparisons in the subsequent analyses.

For the first analysis comparing predictions in the two groups, due to variability in the total number and the frequency of visits between participants, we aimed to create a matched dataset based on visit history from the nMCI and pMCI cohorts to compare the predictions in the two groups with reference to a specific timepoint. We first started with all the progressive patients $p_p \in pMCI$ who got a diagnosis at timepoint t_{n+1}, and, for each patient individually, compiled all previous visits t_m, $m \leq n$ into a vector h_p representing the time of the visits. The entries d_{t_m} of the vector were the number of days until the diagnosis was given, $t_{n+1} - t_m$. For simplicity we also appended $d_{t_{n+1}} = 0$ to h_p, such that for a single patient

$$h_p = \left[d_{t_0},\ d_{t_1},\ \ldots,\ d_{t_n},\ 0\right].$$

Then, for each of the non-progressive patients $p_n \in nMCI$ who didn't have a time of diagnosis (e.g. t_{n+1} is not given) we compiled a set H_p of all possible history vectors h_p by varying which visit was chosen as t_0 and a terminal non-diagnosis timepoint t_{n+1}. Next, we defined a cost-criterion for matching two histories (with an equal number of visits) as the sum of absolute pairwise differences between the vectors,

$$cost(h_1, h_2) = \sum_{m=0}^{n} |d_{t_m}^{h_1} - d_{t_m}^{h_2}|.$$

For each pair of progressive and non-progressive patients (p_p, p_n) this allowed us to calculate a best possible match, given that the stable patient had a total number of visits equal to or larger than the number of visits for the progressive patient:

$$match(p_p, p_s) = \begin{cases} \min\limits_{h \in H_{p_s}} cost\left(h_{p_p}, h\right) & \exists h \in H_{p_s}\left(|h| = \left|h_{p_p}\right|\right) \\ \infty & else \end{cases}.$$

Finally, we compiled the cost of the optimal match from all pairs into a matrix and found the best complete matching by minimizing the total cost across this matrix using the Hungarian algorithm implemented in scipy v1.6.3 [77], such that each patient occurs in at most one pair.

We estimated differences in predictions $\hat{y}$ between the two groups using a linear mixed model. Specifically, we modelled $\hat{y}$ at all timepoints before the terminal timepoint t_{n+1} as a function of age, sex (as controlling variables), years to diagnosis, categorical group membership (nMCI, pMCI), and an interaction between years to diagnosis and group. In addition, we had an independent intercept and slope per participant. The model was fit the formula API of statsmodels v0.13.2 [78] using the formula

$$y \sim age + sex + years\ to\ diagnosis + C(group) + years\ to\ diagnosis{:}\ C(group) + (1 + years\ to\ diagnosis \mid subject)$$

on the matched dataset. A full overview of coefficients and p-values can be found in Supplementary Table 7.

Due to the high dimensionality of the relevance maps, we decomposed them with a principal component analysis (PCA) before the final analyses. To fit the PCA we used the non-linearly registered relevance maps from a randomly selected timepoint for all MCI patients. Before fitting the model, all relevance maps were smoothed with a constant 3x3x3 blurring kernel using the convolution operation from Tensorflow 2.6 to strengthen the signal-to-noise ratio. The PCA was

computed using scikit-learn v1.0.2 [79], retaining 64 components (out of 1137 maximally possible) in a component vector $c = [c_0, c_1, \ldots, c_{63}]$. An axial slice from each of the 64 components visualized in MNI152-space is shown in Supplementary Figure 9.

We fit Cox proportional hazard models using the component vectors as predictors to assess the association between the relevance maps and progression as a function of age. In addition to the components, representing the maps, we controlled for sex in the model. The p-values and coefficient can be found in Supplementary Table 7. To account for covariance between the components and the dementia-prediction $\hat{y}$ (Supplementary Figure 10) we ran an additional model where we divided the patients into ten strata based on $\hat{y}$. Both models were fit using lifelines v0.27.1 [80].

To further explore the prognostic efficacy of our pipeline we set up a predictive analysis for predicting progression at multiple, fixed timepoints a given number of months in the future. For each participant p with visits at timepoints t^p, we denoted the last timepoint with an MCI diagnosis t^p_{neg} and the first timepoint with a dementia diagnosis (if present) t^p_{pos}. Using a fixed set of years into the future, $\gamma \in \{1, 2, 3, 4, 5\}$, we constructed a target variable $z_\gamma(t^p)$ such that

$$z_\gamma(t^p) = \begin{cases} 1 & t^p + \gamma \geq t^p_{pos} \\ 0 & t^p + \gamma \leq t^p_{neg} \\ NA & else \end{cases}$$

where the NAs allow for exclusion of all patients where the status at timepoint $t^p + \gamma$ is unknown. For each γ we constructed the target vector z_γ across all timepoints for all participants with $z_\gamma \neq NA$ and split the constituent patients p into five folds stratified on z_γ, sex and age, such that all timepoints from a participant resided in the same fold. Using these folds, we fit logistic regression models to predict z_γ with an l_1-penalty in a nested cross-validation loop, allowing us to both tune the regularization parameter λ and have out-of-sample predictions for all participants. For eligible participants we used all timepoints for training the models, but

during testing we sampled a random timepoint per participant to ensure independence between datapoints in the final evaluation. For each γ we fit three models: a baseline model

$$\mathcal{M}_{base} \coloneqq z_\gamma \sim age_{t^p} + sex + age_{t^p} \times sex$$

to assess the bias in the dataset with respect to age at the given timepoint t^p and sex, a model using the prediction $\hat{y}_{t^p}$ from the dementia classifier at t^p as a predictor

$$\mathcal{M}_{pred} \coloneqq z_\gamma \sim age_{t^p} + sex + age_{t^p} \times sex + \hat{y}_{t^p} + age_{t^p} \times \hat{y}_{t^p}$$

and a model including the relevance maps from t^p, represented by the component vector c_{t^p},

$$\mathcal{M}_{comp} \coloneqq z_\gamma \sim age_{t^p} + sex + age_{t^p} \times sex + \hat{y}_{t^p} + age_{t^p} \times \hat{y}_{t^p} + c_{t^p}.$$

All models were fit and tuned using the LogisticRegressionCV interface of sklearn v1.0.2 [79]. We compared models by measuring the mean AUC across the five folds (Supplementary Table 8). To evaluate clinical applicability we also report accuracy, positive predictive value, sensitivity, and specificity (Table 2). To determine whether the more complex models represented significant improvements we employed a Wilcoxon signed-rank test from scipy v1.9.3 [77] to do pairwise comparisons between $\mathcal{M}_{base}$ and $\mathcal{M}_{pred}$, and $\mathcal{M}_{pred}$, and $\mathcal{M}_{comp}$ on results from the five out-of-sample AUCs independently.

To assess whether the relevance maps were associated with specific cognitive functions we associated aspects of them with performance on various cognitive tests. We first extracted test results from seven neuropsychological batteries which spanned all ADNI waves and contained high-level summary scores from the ADNI website (Supplementary Table 9). We then manually extracted 17 summary scores spanning different, but overlapping, cognitive domains (Supplementary Table 10). The component vectors c were used as proxies for the relevance maps, where each c_i represented a template for localization of pathology. We matched 2402 component vectors with test results from 733 MCI patients, forming a basis for the comparison. We then calculated the univariate association between cognitive performance according to each

of the 17 with each of the dimensions $c_i \in c$, while including age and sex as covariates for correction. To isolate the effect of the localization we also corrected for dementia-prediction, $\hat{y}$. When a patient had multiple potential matches, a random timepoint was selected, and the final number of datapoints used in the analyses varied from 518 to 675. Correction for multiple testing was done with the Benjamini-Hochberg procedure. The coefficients and p-values of all correlations are reported in Supplementary Table 13. To ensure the associations were not confounded by collinearities between c and $\hat{y}$, we also performed an equivalent analysis without correction to observe whether the sign of the coefficients changed (Supplementary Table 14).

References

1. Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci* **20**, 365–377 (2017).
2. Bethlehem, R. a. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
3. Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
4. Arbabshirani, M. R., Plis, S., Sui, J. & Calhoun, V. D. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage* **145**, 137–165 (2017).
5. Sui, J., Jiang, R., Bustillo, J. & Calhoun, V. Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biological Psychiatry* **88**, 818–828 (2020).
6. Davatzikos, C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* **23**, 17–20 (2004).
7. Westlin, C. *et al.* Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences* **27**, 246–257 (2023).
8. Bzdok, D. & Ioannidis, J. P. A. Exploration, Inference, and Prediction in Neuroscience and Biomedicine. *Trends in Neurosciences* **42**, 251–262 (2019).
9. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
10. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* **1**, 206–215 (2019).
11. Gauthier S, Webster C, Servaes S, Morais JA, Rosa-Neto P. World Alzheimer Report 2022 – Life after diagnosis: Navigating treatment, care and support. (2022).
12. Nichols, E. *et al.* Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* **18**, 88–106 (2019).
13. Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1204–1222 (2020).

14. World Health Organization. Global status report on the public health response to dementia. (2021).
15. Nichols, E. *et al.* Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health* **7**, e105–e125 (2022).
16. Feldman, H. H. *et al.* Diagnosis and treatment of dementia: 2. Diagnosis. *CMAJ* **178**, 825–836 (2008).
17. Karantzoulis, S. & Galvin, J. E. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev Neurother* **11**, 1579–1591 (2011).
18. Echávarri, C. *et al.* Co-occurrence of Different Pathologies in Dementia: Implications for Dementia Diagnosis. *Journal of Alzheimer's Disease* **30**, 909–917 (2012).
19. Schneider, J. A. Neuropathology of Dementia Disorders. *CONTINUUM: Lifelong Learning in Neurology* **28**, 834 (2022).
20. Ryan, J., Fransquet, P., Wrigglesworth, J. & Lacaze, P. Phenotypic Heterogeneity in Dementia: A Challenge for Epidemiology and Biomarker Studies. *Front Public Health* **6**, 181 (2018).
21. Ikram, M. A. *et al.* Brain tissue volumes in relation to cognitive function and risk of dementia. *Neurobiology of Aging* **31**, 378–386 (2010).
22. McDonald, C. R. *et al.* Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. *Neurobiology of Aging* **33**, 242–253 (2012).
23. Ferreira, D., Nordberg, A. & Westman, E. Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology* **94**, 436–448 (2020).
24. Verdi, S., Marquand, A. F., Schott, J. M. & Cole, J. H. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. *Brain* **144**, 2946–2953 (2021).
25. Rasmussen, J. & Langerman, H. Alzheimer's Disease – Why We Need Early Diagnosis. *Degener Neurol Neuromuscul Dis* **9**, 123–130 (2019).
26. Robinson, L., Tang, E. & Taylor, J.-P. Dementia: timely diagnosis and early intervention. *BMJ* **350**, h3029 (2015).

27. Lu, B. *et al.* A practical Alzheimer's disease classifier via brain imaging-based deep learning on 85,721 samples. *Journal of Big Data* **9**, 101 (2022).
28. Mirzaei, G. & Adeli, H. Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia. *Biomedical Signal Processing and Control* **72**, 103293 (2022).
29. Mirabnahrazam, G. *et al.* Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. *Neurobiology of Aging* **121**, 139–156 (2023).
30. Castellazzi, G. *et al.* A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features. *Frontiers in Neuroinformatics* **14**, (2020).
31. Yao, A. D., Cheng, D. L., Pan, I. & Kitamura, F. Deep Learning in Neuroradiology: A Systematic Review of Current Algorithms and Approaches for the New Wave of Imaging Technology. *Radiology: Artificial Intelligence* **2**, e190026 (2020).
32. Kundu, S. AI in medicine must be explainable. *Nat Med* **27**, 1328–1328 (2021).
33. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. & Müller, K.-R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. (Springer Nature, 2019).
34. Barredo Arrieta, A. *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **58**, 82–115 (2020).
35. Samek, W. & Müller, K.-R. Towards Explainable Artificial Intelligence. in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (eds. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. & Müller, K.-R.) 5–22 (Springer International Publishing, 2019). doi:10.1007/978-3-030-28954-6_1.
36. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. Preprint at https://doi.org/10.48550/arXiv.1312.6034 (2014).
37. Bach, S. *et al.* On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE* **10**, e0130140 (2015).
38. Martin, S. A., Townend, F. J., Barkhof, F. & Cole, J. H. Interpretable machine learning for dementia: A systematic review. *Alzheimer's & Dementia* **n/a**, (2023).

39. Böhle, M., Eitel, F., Weygandt, M. & Ritter, K. Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer's Disease Classification. *Frontiers in Aging Neuroscience* **11**, (2019).
40. Wang, D. *et al.* Deep neural network heatmaps capture Alzheimer's disease patterns reported in a large meta-analysis of neuroimaging studies. Preprint at https://doi.org/10.48550/arXiv.2207.11352 (2022).
41. Kohlbrenner, M. *et al.* Towards Best Practice in Explaining Neural Network Decisions with LRP. in *2020 International Joint Conference on Neural Networks (IJCNN)* 1–7 (2020). doi:10.1109/IJCNN48605.2020.9206975.
42. Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F. & Fox, P. T. Activation Likelihood Estimation meta-analysis revisited. *Neuroimage* **59**, 2349–2361 (2012).
43. Briechle, K. & Hanebeck, U. D. Template matching using fast normalized cross correlation. in (eds. Casasent, D. P. & Chao, T.-H.) 95–102 (2001). doi:10.1117/12.421129.
44. Eitel, F. & Ritter, K. Testing the Robustness of Attribution Methods for Convolutional Neural Networks in MRI-Based Alzheimer's Disease Classification. in *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support* (eds. Suzuki, K. et al.) 3–11 (Springer International Publishing, 2019). doi:10.1007/978-3-030-33850-3_1.
45. Erasmus, A., Brunet, T. D. P. & Fisher, E. What is Interpretability? *Philos. Technol.* **34**, 833–862 (2021).
46. Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* **3**, e745–e750 (2021).
47. Amann, J. *et al.* To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health* **1**, e0000016 (2022).
48. Adebayo, J. *et al.* Sanity Checks for Saliency Maps. *arXiv:1810.03292 [cs, stat]* (2020).
49. Kindermans, P.-J. *et al.* The (Un)reliability of Saliency Methods. in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (eds. Samek, W., Montavon, G.,

Vedaldi, A., Hansen, L. K. & Müller, K.-R.) 267–280 (Springer International Publishing, 2019). doi:10.1007/978-3-030-28954-6_14.

50. Nie, X. *et al.* Subregional Structural Alterations in Hippocampus and Nucleus Accumbens Correlate with the Clinical Impairment in Patients with Alzheimer's Disease Clinical Spectrum: Parallel Combining Volume and Vertex-Based Approach. *Frontiers in Neurology* **8**, (2017).
51. Poulin, S. P., Dautoff, R., Morris, J. C., Barrett, L. F. & Dickerson, B. C. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Res* **194**, 7–13 (2011).
52. Van Hoesen, G. W., Augustinack, J. C., Dierking, J., Redman, S. J. & Thangavel, R. The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates. *Ann N Y Acad Sci* **911**, 254–274 (2000).
53. Visser, P. J. *et al.* Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. *J Neurol* **246**, 477–485 (1999).
54. Echávarri, C. *et al.* Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. *Brain Struct Funct* **215**, 265–271 (2011).
55. Dickerson, B. C. *et al.* Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology* **76**, 1395–1402 (2011).
56. Rafii, M. S. & Aisen, P. S. Detection and treatment of Alzheimer's disease in its preclinical stage. *Nat Aging* **3**, 520–531 (2023).
57. Frisoni, G. B. *et al.* Dementia prevention in memory clinics: recommendations from the European task force for brain health services. *The Lancet Regional Health – Europe* **26**, (2023).
58. de Vugt, M. E. & Verhey, F. R. J. The impact of early dementia diagnosis and intervention on informal caregivers. *Progress in Neurobiology* **110**, 54–62 (2013).
59. Leonardsen, E. H. *et al.* Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage* **256**, 119210 (2022).
60. Mårtensson, G. *et al.* The reliability of a deep learning model in clinical out-of-distribution MRI data: A multicohort study. *Medical Image Analysis* **66**, 101714 (2020).

61. Herzog, C. On the Ethical and Epistemological Utility of Explicable AI in Medicine. *Philos. Technol.* **35**, 50 (2022).
62. Petersen, R. C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209 (2010).
63. Gong, W., Beckmann, C. F., Vedaldi, A., Smith, S. M. & Peng, H. Optimising a Simple Fully Convolutional Network for Accurate Brain Age Prediction in the PAC 2019 Challenge. *Frontiers in Psychiatry* **12**, (2021).
64. Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. & Smith, S. M. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis* **68**, 101871 (2021).
65. Ségonne, F. *et al.* A hybrid approach to the skull stripping problem in MRI. *Neuroimage* **22**, 1060–1075 (2004).
66. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* **5**, 143–156 (2001).
67. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
68. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 19 (2015).
69. Chollet, F. & others. Keras. https://github.com/fchollet/keras (2015).
70. Smith, L. N. Cyclical Learning Rates for Training Neural Networks. *arXiv:1506.01186 [cs]* (2017).
71. Guillemot, M., Heusele, C., Korichi, R., Schnebert, S. & Chen, L. Breaking Batch Normalization for better explainability of Deep Neural Networks through Layer-wise Relevance Propagation. *arXiv:2002.11018 [cs, stat]* (2020).
72. Montavon, G., Binder, A., Lapuschkin, S., Samek, W. & Müller, K.-R. Layer-Wise Relevance Propagation: An Overview. in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (eds. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. & Müller, K.-R.) 193–209 (Springer International Publishing, 2019). doi:10.1007/978-3-030-28954-6_10.
73. Brett, M. *et al.* nipy/nibabel: 3.2.2. (2022) doi:10.5281/zenodo.6617121.

74. Laird, A. R., Lancaster, J. L. & Fox, P. T. BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* **3**, 65–78 (2005).
75. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
76. Samek, W., Binder, A., Montavon, G., Lapuschkin, S. & Müller, K.-R. Evaluating the Visualization of What a Deep Neural Network Has Learned. *IEEE Transactions on Neural Networks and Learning Systems* **28**, 2660–2673 (2017).
77. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
78. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. in 92–96 (2010). doi:10.25080/Majora-92bf1922-011.
79. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
80. Davidson-Pilon, C. lifelines: survival analysis in Python. *Journal of Open Source Software* **4**, 1317 (2019).

Author contributions:

Conceptualization: EHL, TW, LTW, YW. Data curation: KP, EW, GS. Formal analysis: EHL. Funding acquisition: OAA, YW. Investigation: EHL, JMR, DVP, TK, AM, OAA, TW, LTW, YW. Methodology: EHL, EG, ND, TS, ØS, TW, LTW, YW. Project administration: GS, OAA, LTW, YW. Software: EHL. Supervision: TW, LTW, YW. Validation: EHL. Visualization: EHL. Writing – original draft: EHL, TW, LTW, YW. Writing – review & editing: KP, EG, ND, TS, JMR, DVP, ØS, TK, EW, AM, GS, OAA.

Acknowledgements

This work was funded by the UiO:LifeScience Convergence Environment (project: 4MENT), the Research Council of Norway (302854), and the European Research Council under the European Union's Horizon 2020 research and Innovation program (802998). The Southern and Eastern Norway Regional Health Authority supported the study through funding for Karin Persson but was not involved in conducting the study or in preparation of the manuscript. We thank the computational resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway – with project no. (nn9769k/ns9769k).

Competing interests:

KP report work with Roche BN29553 and Novo Nordisk NN6535-4730 trials; All other authors declare that they have no competing interests.

Data availability

The data used in this study were gathered from various sources, an overview including acknowledgements of their respective funding sources is provided in Supplementary Table 1. Among others, data used in the preparation of this article was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI, see adni.loni.usc.edu for further details), the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL, www.aibl.csiro.au) the AddNeuroMed consortium, and MIRIAD (www.nitrc.org/projects/miriad). The investigators within these studies contributed to the design and implementation of the data collection process

but did not participate in the analysis or writing of this report, and this publication is solely the responsibility of the authors.

Code availability:

The trained model and explainable pipeline and the underlying code will be made available at https://github.com/estenhl/pyment-public upon publication. Generic code for generating explanations for 3D CNNs is available at https://github.com/estenhl/keras-explainability.

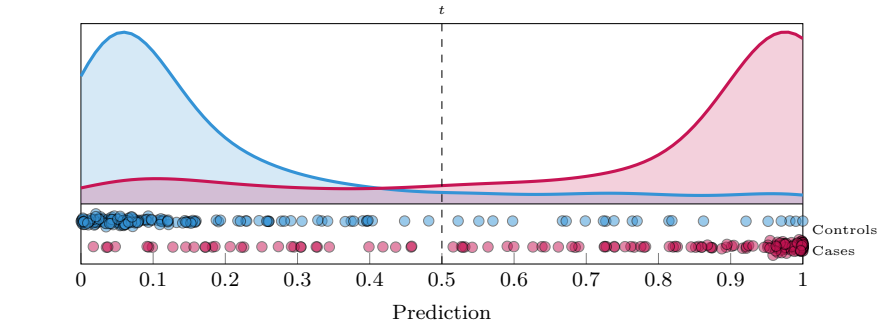
List of supplementary figures

1 Test performance for the dementia classifiers 3
2 Example relevance maps 4
3 Average relevance maps 5
4 Statistical map from GingerALE 6
5 Morphological record for a stable MCI patient 7
6 Morphological record for an improving MCI patient 8
7 Distributions of dementia predictions in MCI subsets 9
8 Variance explained by PCA components 10
9 Visualization of PCA components 11
10 Correlations between PCA components 12
11 Neuropsychological tests and PCA components 13
12 Learning rates for dementia classifiers 14
13 Results from the hyperparameter search 15
14 Region-wise comparison of $\bar{R}_{dementia}$ and G 16

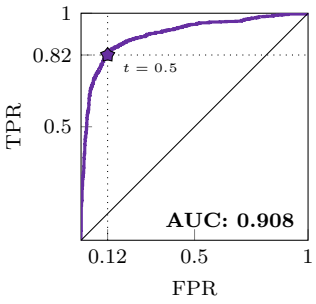
List of supplementary tables

1 Overview of datasets 17
2 Inclusion criteria for participants 18
3 Predictive performance of dementia classifiers by fold 19
4 Predictive performance of dementia classifiers by site 20
5 Overview of composite LRP strategy 21
6 Effects of various covariates on the dementia prediction $\hat{y}$ 22
7 Overview of survival analyses 23
8 Predictive performance of progression models 24
9 Neuropsychological assessments 25

10 Neuropsychological test scores 26
11 Dementia classifier hyperparameters 27
12 Dementia classifier augmentations 28



(a) Predictions

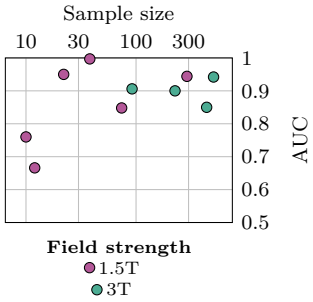


(b) ROC Curve

Observed \ Predicted	0	1
0	754	100
1	157	697

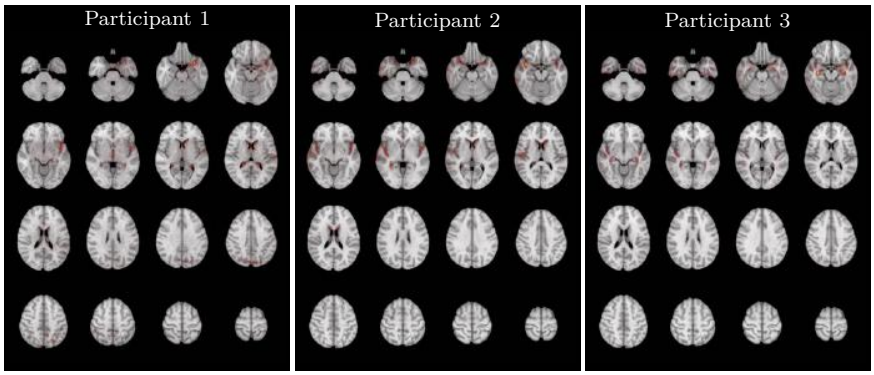
Accuracy: 84.95%

(c) Confusion matrix

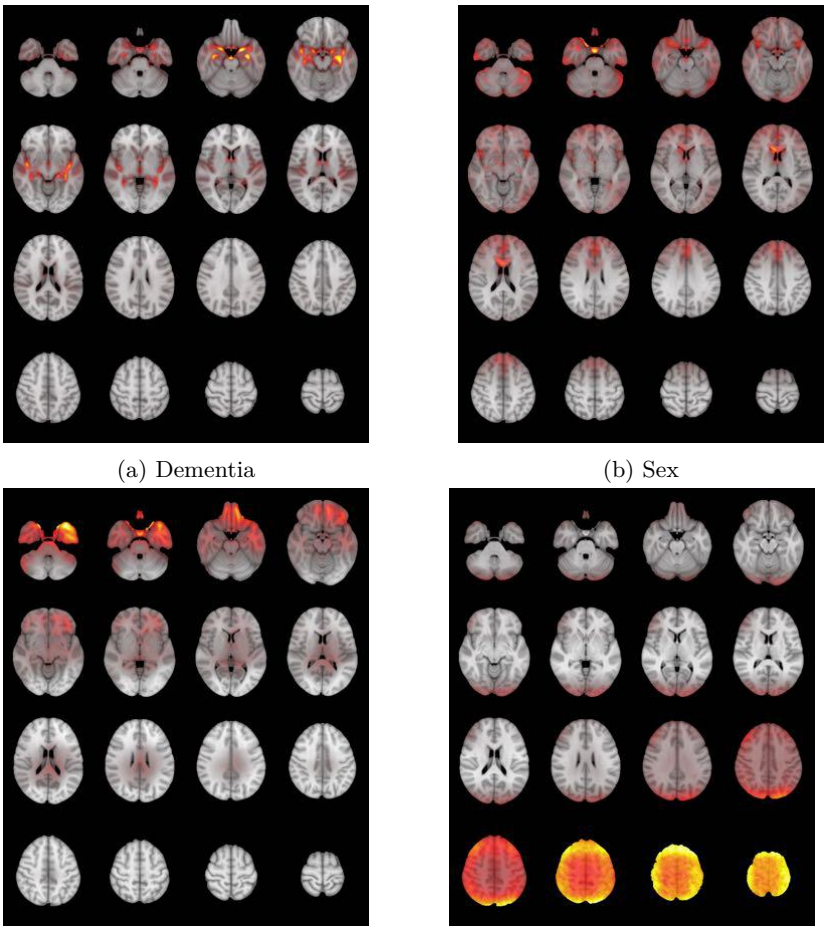


(d) AUC per site

Supplementary Figure 1: Predictive performance of the best performing dementia classifiers, combined across each out-of-sample test fold.



Supplementary Figure 2: Example relevance maps produced by the pipeline for three randomly selected dementia patients.



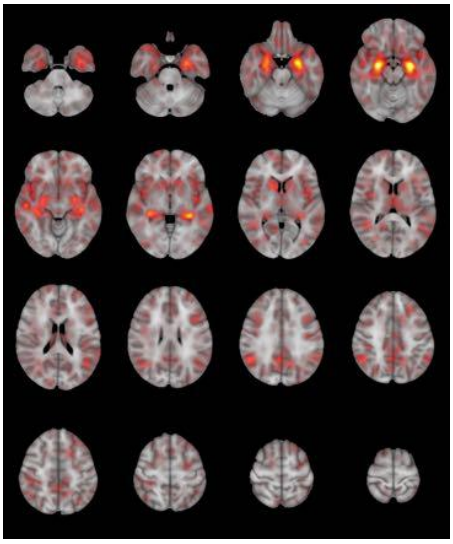
(a) Dementia

(b) Sex

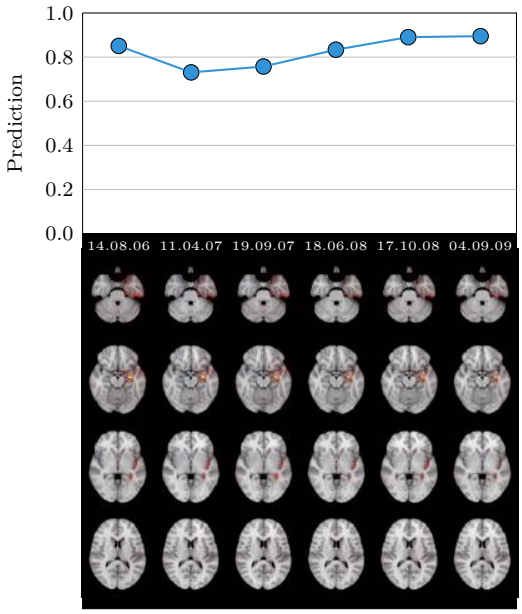
(c) Random weights

(d) Randomized images

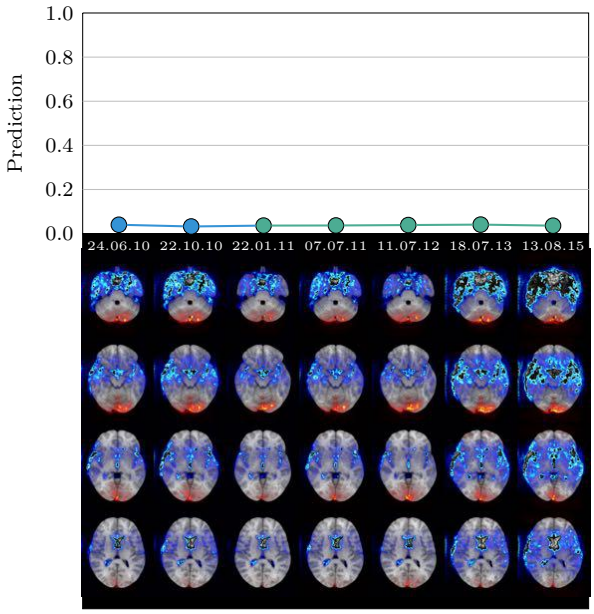
Supplementary Figure 3: Average voxel-wise relevance maps produced for the four pipelines across all correctly predicted dementia patients.



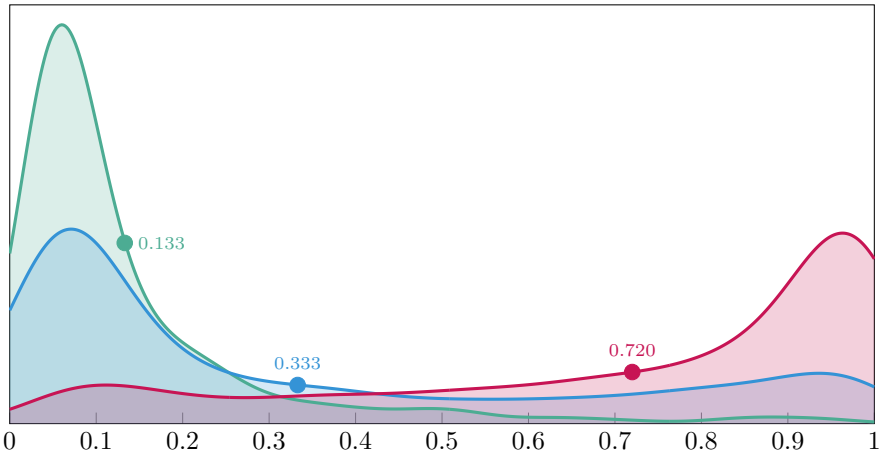
Supplementary Figure 4: Statistical map of dementia-related pathology generated by a GingerALE meta-analysis of 124 articles.



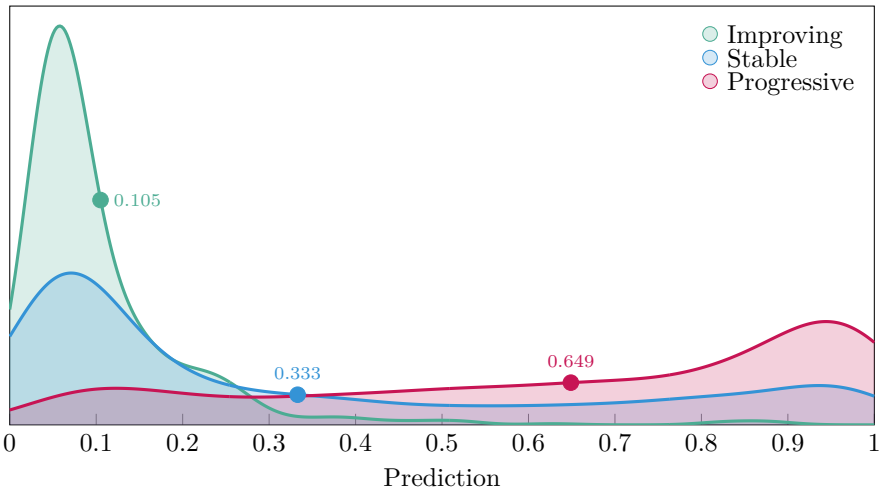
Supplementary Figure 5: Morphological record for a randomly selected stable MCI patient across six visits.



Supplementary Figure 6: Morphological record of a randomly selected improving MCI patient across seven visits.

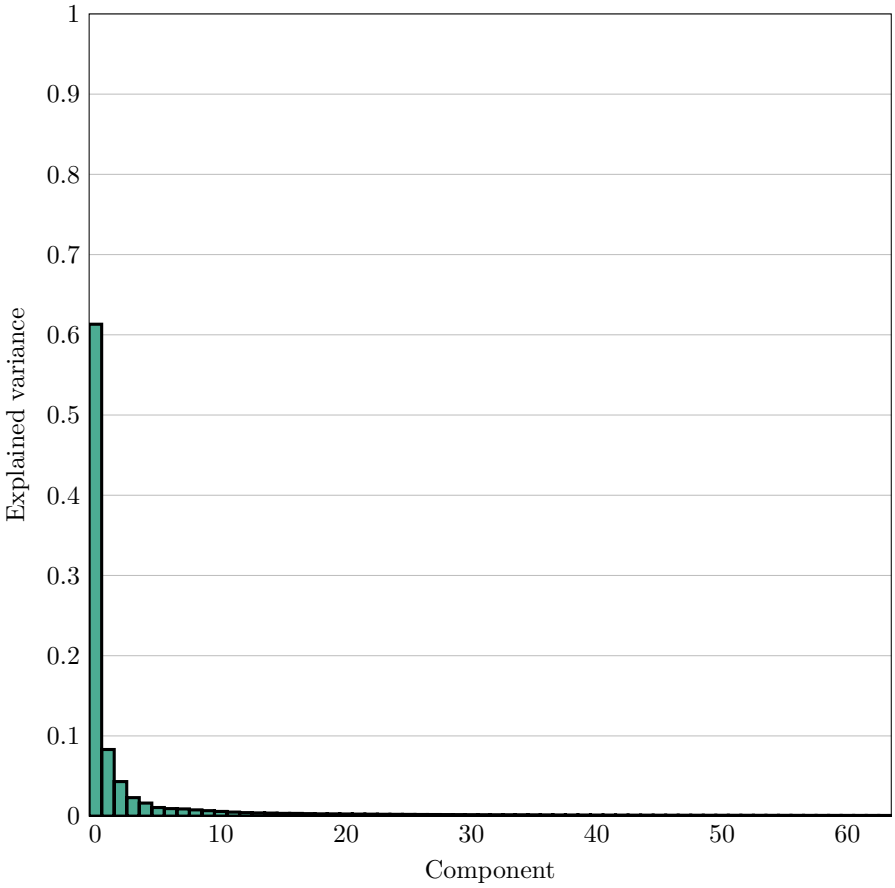


(a) All timepoints.

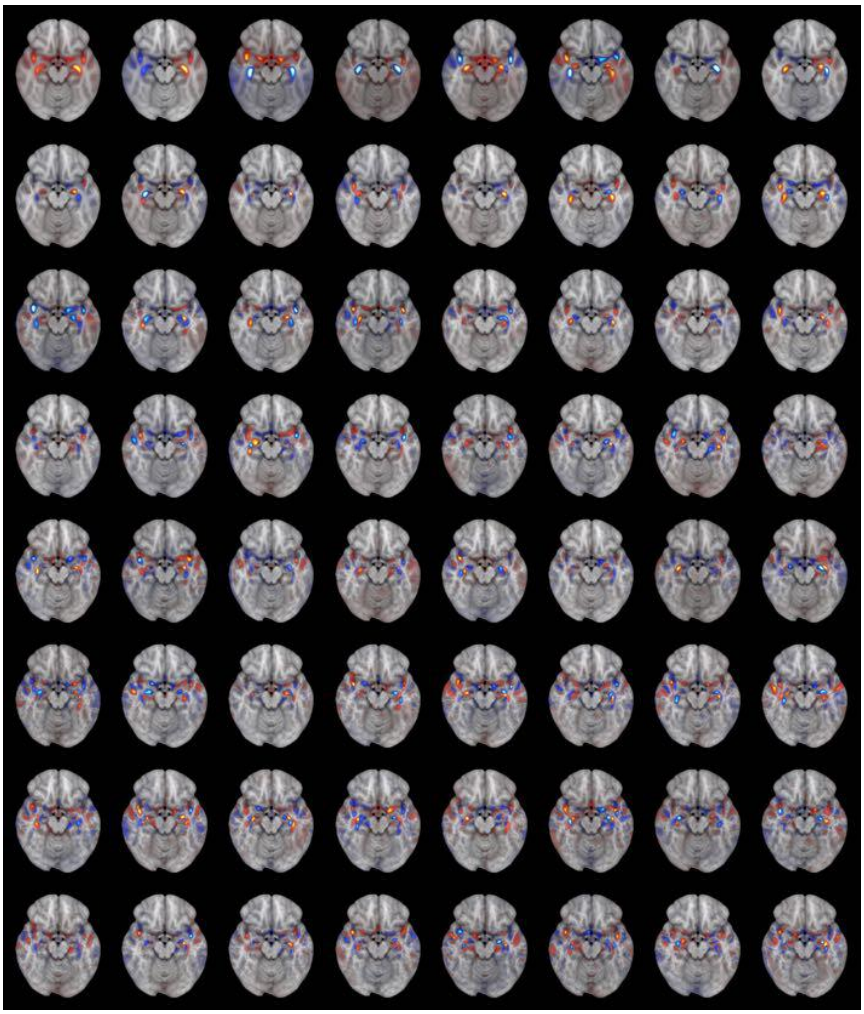


(b) Only timepoints when the patient got an MCI diagnosis.

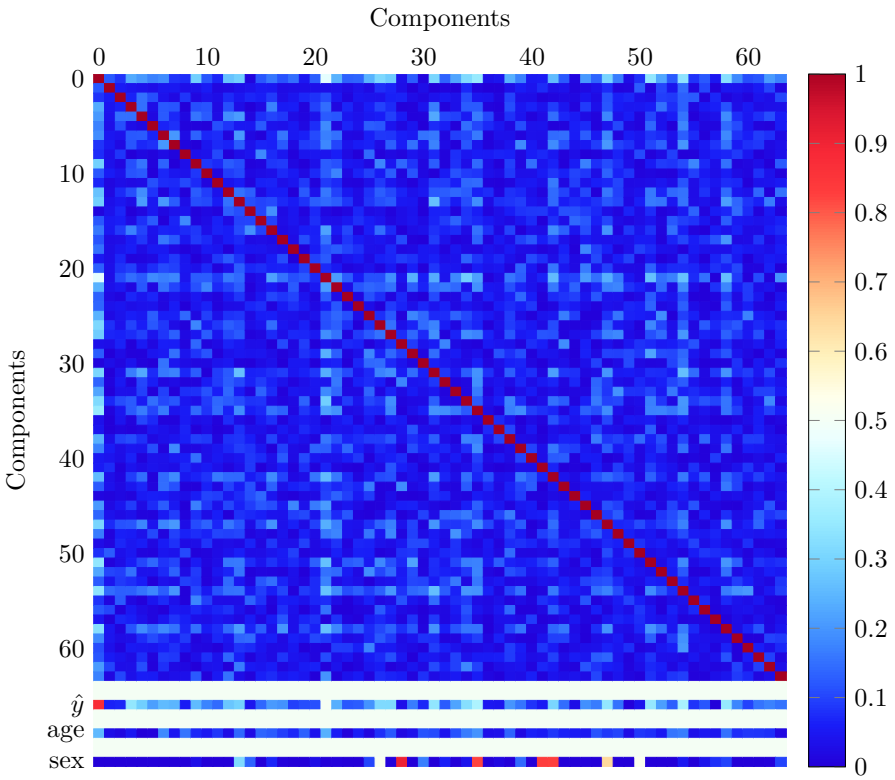
Supplementary Figure 7: Distribution of predictions from the dementia classifier in the three MCI subsets. The marks indicate the mean prediction in each subset.



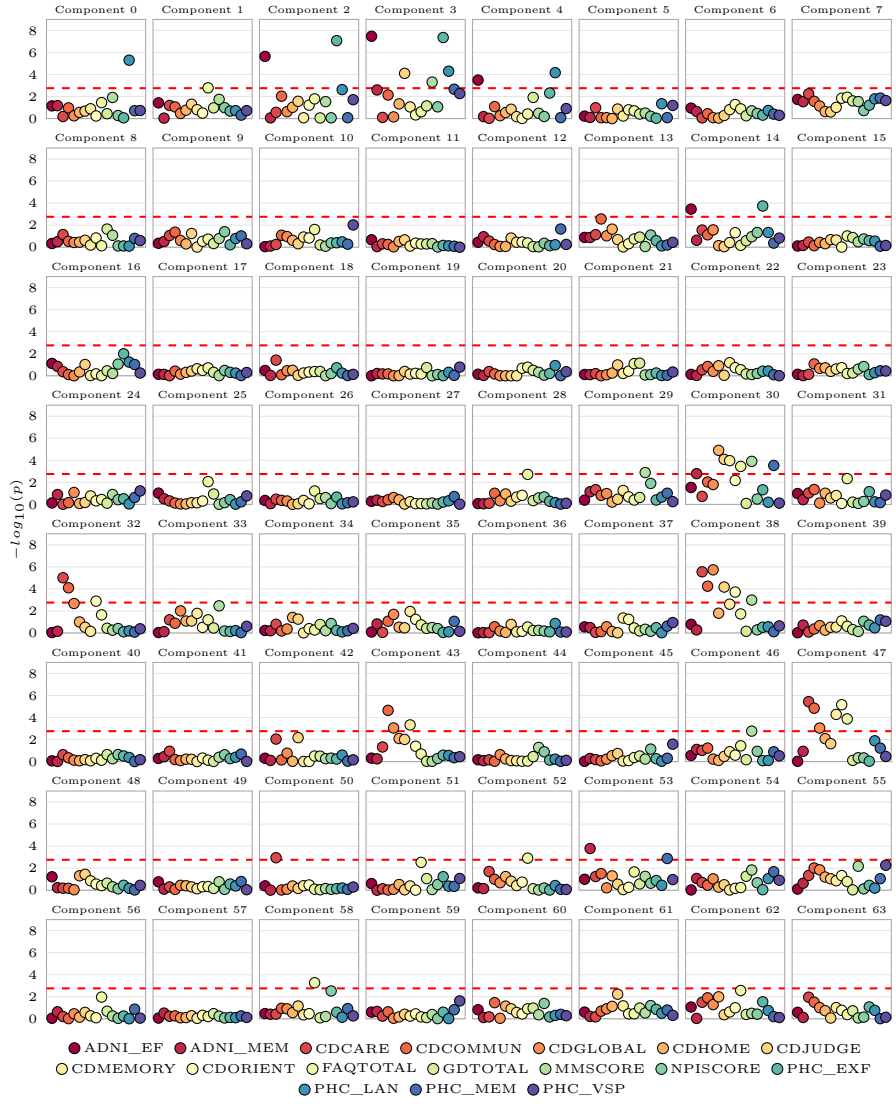
Supplementary Figure 8: Explained variance for each of the 64 selected PCA components.



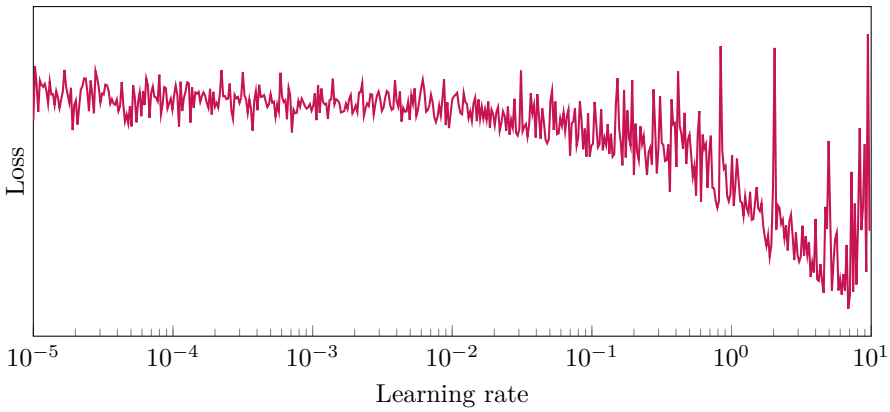
Supplementary Figure 9: A single axial slice from each of the 64 principal components plotted in brain space.



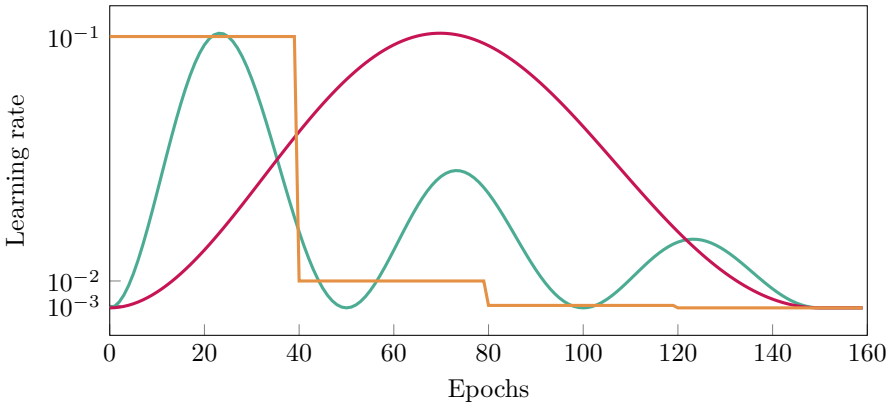
Supplementary Figure 10: Magnitudes of the correlations within the 64 principal components from the PCA, and between the components and prediction, age and sex. Correlations were computed in the dataset used for the exploratory MCI analyses.



Supplementary Figure 11: P-values of the associations between the neuropsychological scores and each of the 64 components.

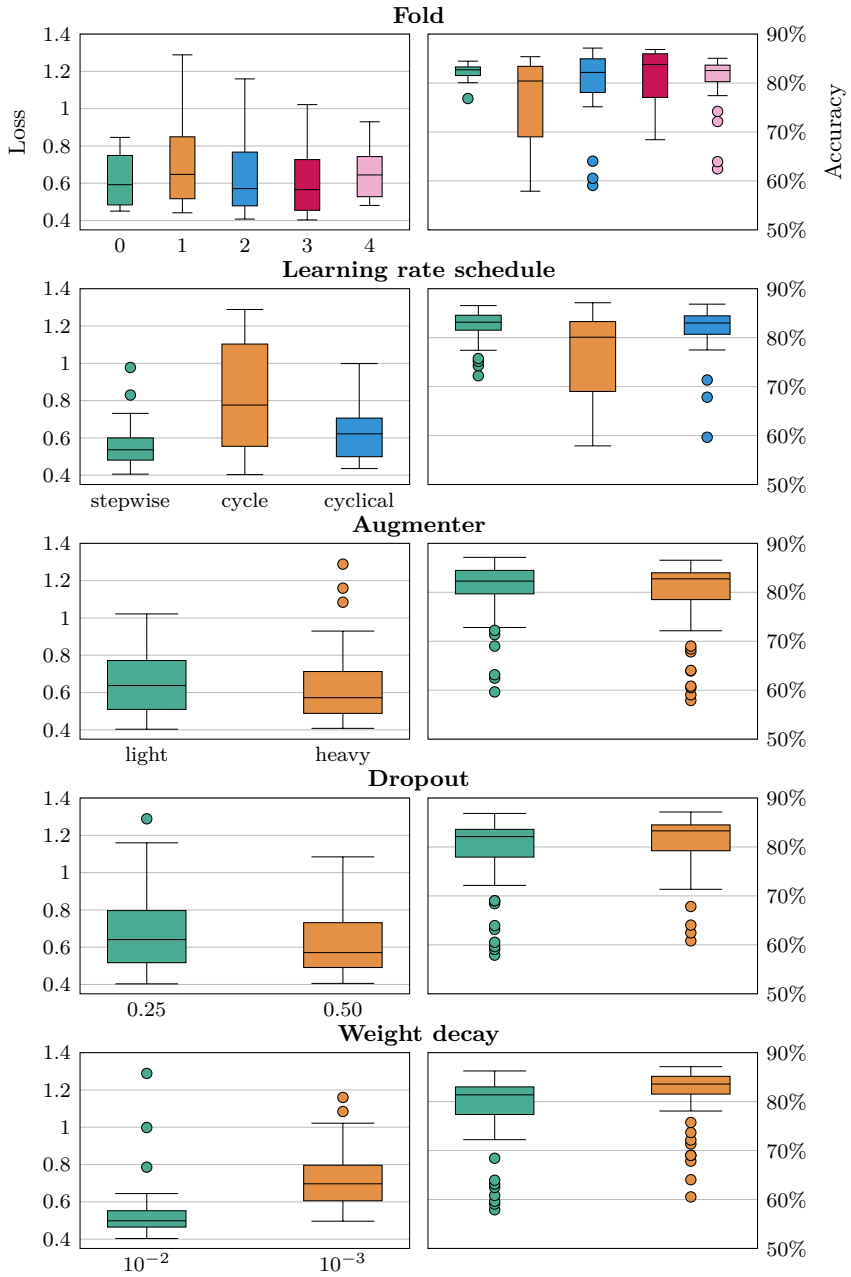


(a) Learning rate sweep

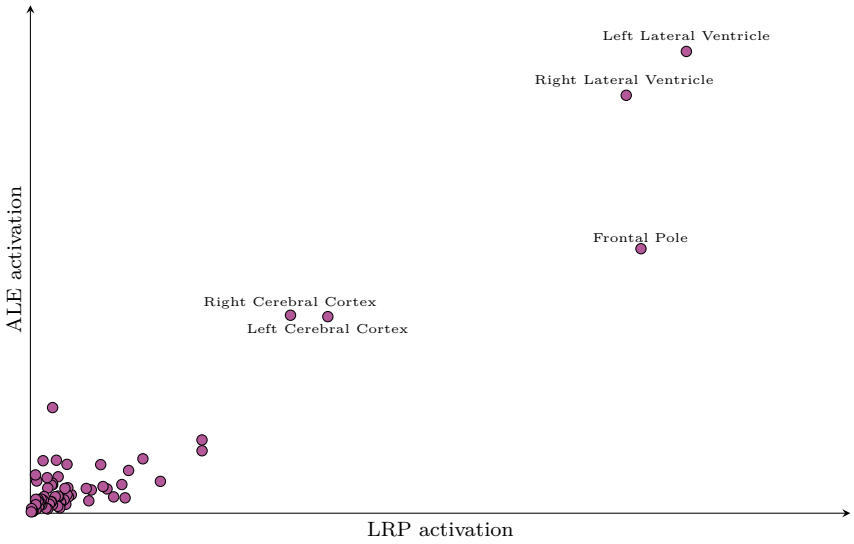


(b) The three different learning rate schedules

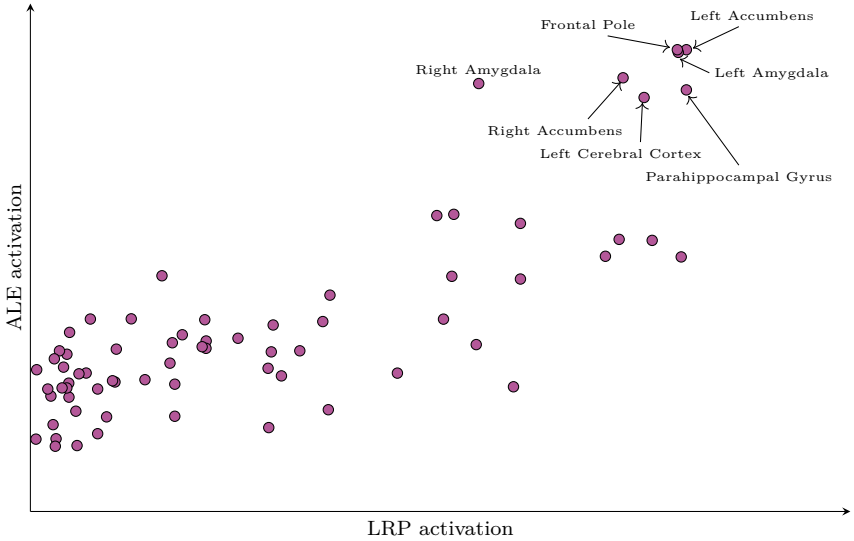
Supplementary Figure 12: Information about the learning rates used when training the dementia classifiers. (a) shows the results of the learning rate sweep, determining the bounds of the learning rate. (b) shows the three different learning rate schedules we employed.



Supplementary Figure 13: Results achieved in the validation folds for the training runs of the dementia classifiers, grouped by each possible setting of each individual hyperparameter.



(a) Total activation per region



(b) Maximum activation per region

Supplementary Figure 14: Region-wise comparison of activation in the average map $\bar{R}_{dementia}$ produced by $LRP_{dementia}$ and G produced by the GingerALE meta-analysis, using different aggregation methods.

Cohort	Source	Comment	References
AddNeuroMed	Authors	AddNeuroMed consortium was led by Simon Lovestone, Bruno Vellas, Patrizia Mecocci, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen. Their work was supported by InnoMed (Innovative Medicines in Europe), an integrated project funded by the European Union of the Sixth Framework program priority (FP6-2004- LIFESCIHEALTH-5)	[1, 2]
Alzheimer's Disease Neuroimaging Initiative	adni.loni.usc.edu	Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.	[3, 4, 5]
Australian Imaging Biomarkers and Lifestyle flagship study of ageing	http://www.aibl.csiro.au/	Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au	[6]
Demgen	Authors	Supported by the Norwegian National Advisory Unit on Aging and Health	[7, 8]
MIRIAD	https://www.nitrc.org/projects/miriad/	Data used in the preparation of this article were obtained from the MIRIAD database (http://miriad.drc.ion.ucl.ac.uk). The MIRIAD investigators did not participate in analysis or writing of this report. The MIRIAD dataset is made available through the support of the UK Alzheimer's Society (Grant RF116). The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC).	[9]
OASIS3	https://www.oasis-brains.org/	Data were provided by OASIS-3: Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly	[10]
StrokeMRI	Authors	Supported by the Research Council of Norway (249795, 248238), the South-Eastern Norway Regional Health Authority (2014097, 2015044, 2015073, 2016083), and the Norwegian ExtraFoundation for Health and Rehabilitation (2015/FO5146)	[11]
TOP	Authors	Supported by several grants from the Research Council of Norway, and the South-Eastern Norway Regional Health Authority	[12, 13, 14]

Supplementary Table 1: An overview of the datasets used in the present study.

Dataset	Controls	Patients
AddNeuroMed	MMSE $\geq$ 24	MMSE $<$ 19
ADNI	DX = CN	DX = AD
AIBL	DX = DXNORM	DX $\in$ {DXAD, DXOTHDEM}
Demgen	-	DX $\in$ {AD, OtherDem, UnspecDem, VaD}
MIRIAD	Group = Control	Group = AD
OASIS3	NORMCOG = 1	NORMCOG = 0 & DEMENTED = 1
StrokeMRI	Group = Control	-
TOP	diagnosis = CTRL	-

Supplementary Table 2: Criteria for inclusion in the case-control groups. For replicability, variable and category names are kept as they were originally used in the originating dataset.

	Validation		Test	
Fold	**AUC**	**Accuracy**	**AUC**	**Accuracy**
0	0.914	84.16%	0.917	83.04
1	0.929	84.79%	0.920	86.25
2	0.925	85.96%	0.915	87.13
3	0.929	87.42%	0.904	84.45
4	0.904	84.16%	0.905	83.87
	0.918	**85.30**	**0.908**	**84.95**

Supplementary Table 3: Predictive performance of the dementia classifiers split up into individual folds. The fold number denotes which fold was used as the test fold for the given run.

Sites	Size	AUC	Accuracy	Confusion Matrix	
ADNI 3.0T	506	0.942	90.11	237	16
				34	219
OASIS 3.0T	438	0.850	79.45	185	34
				56	163
ADNI 1.5T	290	0.944	87.93	125	20
				15	130
Oslo GE750	226	0.900	84.07	102	11
				25	88
AIBL Site 1	92	0.906	84.78	43	3
				11	35
ANM GE	74	0.848	72.97	26	11
				9	28
Miriad 1.5T	38	0.997	97.36	19	0
				1	18
AIBL Site 2	22	0.950	86.36	11	0
				3	8
OASIS 1.5T	12	0.666	58.33	3	3
				2	4
ANM Picker	10	0.76	70.00	3	2
				1	4

Supplementary Table 4: Predictive performance of the dementia classifiers split up into individual proxy sites.

Layer	**LRP Strategy**	**Parameters**
Conv3D	flat	$\{\flat = \text{True}\}$
Conv3D	flat	$\{\flat = \text{True}\}$
Conv3D	alpha-beta	$\{\alpha = 1, \beta = 0\}$
Conv3D	alpha-beta	$\{\alpha = 1, \beta = 0\}$
Conv3D	alpha-beta	$\{\alpha = 1, \beta = 0\}$
Conv3D	alpha-beta	$\{\alpha = 1, \beta = 0\}$
Dense	epsilon	$\{\epsilon = 0.25\}$

Supplementary Table 5: Composite LRP strategy that was used for generating the explanations.

Variable	**Coefficient**	**p**
intercept	-0.373	1.94×10^{-3}
sex[M]	-0.084	7.29×10^{-4}
group[pMCI]	0.473	6.05×10^{-71}
age	0.009	8.76×10^{-10}
years_to_diagnosis	0.013	3.92×10^{-3}
years_to_diagnosis:group[pMCI]	0.050	8.14×10^{-17}

Supplementary Table 6: Coefficients and p-values from the linear mixed model modelling dementia prediction $\hat{y}$ as a function of various variables, including membership in the progressive (pMCI) or non-progressive group.

	Uncorrected		Corrected for $\hat{y}$	
Component	**p**	β	**p**	β
0	$\mathbf{1.60 \times 10^{-66}}$	0.683	**0.003**	0.18
1	0.061	−0.043	**0.001**	−0.07
2	$\mathbf{4.41 \times 10^{-26}}$	−0.248	$\mathbf{1.74 \times 10^{-20}}$	−0.22
3	$\mathbf{2.32 \times 10^{-20}}$	0.222	$\mathbf{9.49 \times 10^{-10}}$	0.16
4	$\mathbf{6.61 \times 10^{-14}}$	0.173	$\mathbf{3.91 \times 10^{-09}}$	0.14
5	0.117	−0.032	0.525	−0.01
6	$\mathbf{1.22 \times 10^{-04}}$	−0.083	$\mathbf{2.84 \times 10^{-05}}$	−0.09
7	0.11	−0.036	0.037	0.05
8	$\mathbf{2.90 \times 10^{-04}}$	0.076	$\mathbf{7.93 \times 10^{-04}}$	0.07
9	0.164	0.028	0.766	0.01
10	0.024	0.051	0.049	0.05
11	$\mathbf{5.74 \times 10^{-05}}$	0.082	$\mathbf{1.25 \times 10^{-04}}$	0.08
12	$\mathbf{3.50 \times 10^{-05}}$	−0.09	$\mathbf{2.78 \times 10^{-06}}$	−0.1
13	**0.003**	−0.063	$\mathbf{1.13 \times 10^{-05}}$	−0.1
14	$\mathbf{6.83 \times 10^{-05}}$	0.088	**0.003**	0.07
15	$\mathbf{1.94 \times 10^{-04}}$	−0.081	$\mathbf{1.91 \times 10^{-06}}$	−0.1
16	**0.001**	−0.071	0.016	−0.05
17	$\mathbf{6.44 \times 10^{-05}}$	−0.087	**0.001**	−0.07
18	**0.002**	−0.064	$\mathbf{9.70 \times 10^{-07}}$	−0.1
19	0.316	−0.022	0.152	−0.03
20	0.492	0.016	0.949	0.0
21	0.484	−0.018	0.374	−0.02
22	**0.008**	−0.067	$\mathbf{6.99 \times 10^{-05}}$	−0.1
23	**0.018**	0.049	0.068	0.04
24	$\mathbf{4.03 \times 10^{-06}}$	−0.106	$\mathbf{1.02 \times 10^{-04}}$	−0.09
25	**0.017**	0.057	$\mathbf{5.75 \times 10^{-04}}$	0.08
26	$\mathbf{1.40 \times 10^{-07}}$	−0.11	$\mathbf{3.63 \times 10^{-07}}$	−0.11
27	0.368	0.019	**0.006**	0.06
28	0.546	0.013	0.229	0.03
29	$\mathbf{4.03 \times 10^{-04}}$	−0.075	$\mathbf{2.80 \times 10^{-04}}$	−0.08
30	$\mathbf{3.44 \times 10^{-08}}$	−0.104	$\mathbf{1.72 \times 10^{-15}}$	−0.16
31	0.859	0.004	0.617	0.01
32	0.061	0.04	0.029	0.05
33	**0.003**	−0.063	$\mathbf{5.71 \times 10^{-04}}$	−0.08
34	0.059	0.04	**0.001**	0.07
35	0.316	−0.024	0.792	0.01
36	0.233	0.023	**0.007**	0.05
37	$\mathbf{1.82 \times 10^{-05}}$	−0.09	$\mathbf{5.41 \times 10^{-06}}$	−0.1
38	$\mathbf{7.93 \times 10^{-04}}$	0.075	**0.001**	0.07
39	0.267	−0.022	0.07	−0.04
40	**0.01**	0.052	$\mathbf{3.62 \times 10^{-05}}$	0.08
41	0.159	−0.034	0.08	−0.04
42	$\mathbf{1.87 \times 10^{-07}}$	−0.116	$\mathbf{2.83 \times 10^{-10}}$	−0.14
43	0.044	0.042	0.104	0.03
44	**0.009**	0.058	0.1	0.04
45	0.157	0.032	0.287	0.02
46	0.81	0.005	0.399	−0.02
47	$\mathbf{2.24 \times 10^{-05}}$	0.087	**0.001**	0.07
48	0.067	0.034	**0.009**	0.05
49	0.159	0.028	**0.009**	0.05
50	0.969	0.001	0.196	−0.02
51	**0.004**	0.065	**0.01**	0.06
52	0.691	0.008	0.424	0.02
53	0.805	−0.005	0.941	−0.0
54	0.204	0.032	0.071	0.05
55	0.082	0.033	0.121	0.03
56	0.207	0.028	0.104	0.04
57	0.147	−0.03	0.052	−0.04
58	$\mathbf{2.18 \times 10^{-08}}$	−0.11	$\mathbf{4.09 \times 10^{-09}}$	−0.12
59	0.05	0.035	$\mathbf{7.21 \times 10^{-04}}$	0.06
60	0.412	−0.017	0.173	−0.03
61	0.554	0.011	0.555	0.01
62	$\mathbf{8.24 \times 10^{-04}}$	0.067	0.023	0.05
63	0.336	0.019	**0.006**	0.05

Supplementary Table 7: Coefficient and p-value for each component in the two Partial Hazard analyses. In the corrected analyses (the two rightmost columns), the patients were stratified based on $\hat{y}$, to isolate the effect of the components.

Model	AUC	Balanced accuracy	PPV	Sensitivity	Specificity
$\mathcal{M}_{base}$	0.506±0.039	49.85%±3.7%	0.11±0.1	0.31±0.26	0.69±0.26
$\mathcal{M}_{pred}$	0.666±0.083	61.49%±10.58%	0.21±0.12	0.56±0.31	0.67±0.18
$\mathcal{M}_{comp}$	0.743±0.063	70.1%±4.79%	0.32±0.03	0.68±0.11	0.72±0.02

(a) One year

Model	AUC	Balanced accuracy	PPV	Sensitivity	Specificity
$\mathcal{M}_{base}$	0.474±0.05	49.47%±5.34%	0.33±0.05	0.47±0.06	0.52±0.07
$\mathcal{M}_{pred}$	0.742±0.053	69.92%±5.24%	0.52±0.07	0.75±0.09	0.65±0.08
$\mathcal{M}_{comp}$	0.786±0.019	74.31%±0.99%	0.61±0.04	0.72±0.04	0.76±0.06

(b) Two years

Model	AUC	Balanced accuracy	PPV	Sensitivity	Specificity
$\mathcal{M}_{base}$	0.536±0.032	53.9%±3.78%	0.32±0.26	0.35±0.29	0.73±0.23
$\mathcal{M}_{pred}$	0.797±0.037	75.76%±4.16%	0.75±0.04	0.76±0.09	0.76±0.06
$\mathcal{M}_{comp}$	0.808±0.016	77.17%±2.29%	0.79±0.04	0.72±0.03	0.82±0.05

(c) Three years

Model	AUC	Balanced accuracy	PPV	Sensitivity	Specificity
$\mathcal{M}_{base}$	0.529±0.056	53.01%±3.74%	0.38±0.32	0.35±0.29	0.71±0.24
$\mathcal{M}_{pred}$	0.844±0.041	80.6%±4.09%	0.84±0.04	0.82±0.06	0.79±0.04
$\mathcal{M}_{comp}$	0.867±0.031	80.38%±3.26%	0.87±0.02	0.75±0.08	0.85±0.02

(d) Four years

Model	AUC	Balanced accuracy	PPV	Sensitivity	Specificity
$\mathcal{M}_{base}$	0.515±0.031	51.05%±2.09%	0.14±0.28	0.09±0.18	0.93±0.14
$\mathcal{M}_{pred}$	0.889±0.024	83.61%±1.81%	0.91±0.02	0.83±0.04	0.84±0.03
$\mathcal{M}_{comp}$	0.903±0.034	84.1%±2.52%	0.92±0.02	0.82±0.03	0.86±0.04

(e) Five years

Supplementary Table 8: Predictive performance for each of the progression models at each of the five years following the MRI examination.

Name	Included	Exclusion reason
Alzheimer's Disease Assessment Scale Sub-scores	No	Missing ADNI2, 3, GO
ADSP Phenotype Harmonization Consortium	Yes	
Alzheimer's Disease Assessment Scale	No	Missing description
Clinical Dementia Rating	Yes	
Cognitive Change Index	No	Missing ADNI1, GO
Cogstate Battery	No	Missing ADNI1, GO
Cogstate Brief Battery	No	Missing ADNI1, 2, GO
Digital Cognitive Biomarkers	No	Missing summary scores
Everyday Cognition	No	Missing ADNI1
Financial Capacity Instrument	No	Missing ADNI1, 2, GO
Functional Activities Questionnaire	Yes	
Geriatric Depression Scale	Yes	
Item level data	No	Missing ADNI2, 3, GO
Mini-Mental State Examination	Yes	
Modified Haschinski Ischemia Scale	No	Non-congruent visit codes
Montreal Cognitive Assessment	No	Missing ADNI1
Neuropsychiatric Inventory Questionnaire	Yes	
Neuropsychological Battery	No	Missing explanation
UW - Neuropsych Summary Scores	Yes	

Supplementary Table 9: Neuropsychological assessments used as a basis for finding associations with the components from the relevance maps.

Variable	Source	Description
PHC_MEM	ADSP Phenotype Harmonization Consortium	Harmonized composite memory score
PHC_EXF	ADSP Phenotype Harmonization Consortium	Harmonized composite executive function score
PHC_LAN	ADSP Phenotype Harmonization Consortium	Harmonized composite language score
PHC_VSP	ADSP Phenotypxe Harmonization Consortium	Harmonized composite visuospatial score
CDRMEMORY	Clinical Dementia Rating	Clinical evaluation of impairment related to memory
CDRORIENT	Clinical Dementia Rating	Clinical evaluation of impairment related to orientation
CDRJUDGE	Clinical Dementia Rating	Clinical evaluation of impairment related to judgement and problem solving
CDRCOMMUN	Clinical Dementia Rating	Clinical evaluation of impairment related to community affairs
CDRHOME	Clinical Dementia Rating	Clinical evaluation of impairment related to home and hobbies
CDRCARE	Clinical Dementia Rating	Clinical evaluation of impairment related to personal care
CDRGLOBAL	Clinical Dementia Rating	Total clinical dementia rating
MMSCORE	Mini-Mental State Examination	Overall measure of cognitive impairment
FAQTOTAL	Functional Activities Questionnaire	Measures instrumental activities from daily life
GDTOTAL	Geriatric Depression Scale	Screening for depression in elderly adults
NPISCORE	Neuropsychiatric Inventory Questionnaire	Informant-based assessment of Neuropsychiatric symptoms
ADNI_EF	UW - Neuropsych Summary Scores	Composite score for executive functioning
ADNI_MEM	UW - Neuropsych Summary Scores	Composite score for memory

Supplementary Table 10: Test scores which were correlated against the components from the relevance maps.

Hyperparameter	**Values**
Dropout	{0.25, 0.5}
Weight decay	{1e-2, 1e-3}
Augmenter	{Light, Heavy}
Learning rate	{Stepwise, Cyclical, One-Cycle}

Supplementary Table 11: Hyperparameters that were tuned when training the dementia classifiers.

Augmentation	**Light**	**Heavy**	**Unit**
Sagittal flip	0.5	0.5	Probability
Sagittal shift	[0,5]	[0,5]	Voxels
Coronal shift	[0,5]	[0,5]	Voxels
Axial shift	[0,5]	[0,5]	Voxels
Zoom		[0, 0.05]	Fraction
Rotation		[0, 5]	Degrees
Noise		[0, 0.1]	Fraction
Blur		[0, 0.2]	Fraction
Crop box	[0, 50]	[0, 50]	Voxels

Supplementary Table 12: Augmentations used in the two augmenters. For the augmentations with a range, values were sampled for each image each epoch in real-time during training.

Supplementary references

[1] Yawu Liu et al. "Combination analysis of neuropsychological tests and structural MRI measures in differentiating AD, MCI and control groups–the AddNeuroMed study". In: *Neurobiology of Aging* 32 (July 2011). ISSN: 1558-1497. DOI: 10.1016/j.neurobiolaging.2009.07.008.

[2] S. Lovestone, P. Francis, and K. Strandgaard. "Biomarkers for disease modification trials–the innovative medicines initiative and AddNeuroMed". In: *The Journal of Nutrition, Health & Aging* 11 (2007). ISSN: 1279-7707.

[3] R C. Petersen et al. "Alzheimer's Disease Neuroimaging Initiative (ADNI)". In: *Neurology* 74 (Jan. 19, 2010). ISSN: 0028-3878. DOI: 10.1212/WNL.0b013e3181cb3e25.

[4] Michael W. Weiner et al. "The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception". In: *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 8 (Feb. 2012). ISSN: 1552-5279. DOI: 10.1016/j.jalz.2011.09.172.

[5] Michael W. Weiner et al. "The Alzheimer's Disease Neuroimaging Initiative 3: continued innovation for clinical trial improvement". In: *Alzheimer's & dementia : the journal of the Alzheimer's Association* 13 (May 2017). ISSN: 1552-5260. DOI: 10.1016/j.jalz.2016.10.006.

[6] Kathryn A. Ellis et al. "The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease". In: *International Psychogeriatrics* 21 (Aug. 2009). ISSN: 1041-6102. DOI: 10.1017/S1041610209009405.

[7] Aldo Córdova-Palomera et al. "Disrupted global metastability and static and dynamic brain connectivity across individuals in the Alzheimer's dis-

ease continuum". In: *Scientific Reports* 7 (Jan. 11, 2017). ISSN: 2045-2322. DOI: 10.1038/srep40268.

[8] Nhat Trung Doan et al. "Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum". In: *Scientific Reports* 7 (Mar. 24, 2017). ISSN: 2045-2322. DOI: 10.1038/srep45131.

[9] Ian B. Malone et al. "MIRIAD–Public release of a multiple time point Alzheimer's MR imaging dataset". In: *NeuroImage* 70 (Apr. 15, 2013). ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2012.12.044.

[10] Pamela J. LaMontagne et al. *OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease*. Dec. 15, 2019. DOI: 10.1101/2019.12.13.19014902.

[11] Erlend S. Dørum et al. "Age-related differences in brain network activation and co-activation during multiple object tracking". In: *Brain and Behavior* 6 (Nov. 2016). ISSN: 2162-3279. DOI: 10.1002/brb3.533.

[12] Lars M. Rimol et al. "Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder". In: *Biological Psychiatry* 68 (July 1, 2010). ISSN: 1873-2402. DOI: 10.1016/j.biopsych.2010.03.036.

[13] Siren Tønnesen et al. "White matter aberrations and age-related trajectories in patients with schizophrenia and bipolar disorder revealed by diffusion tensor imaging". In: *Scientific Reports* 8 (Sept. 20, 2018). ISSN: 2045-2322. DOI: 10.1038/s41598-018-32355-9.

[14] Nhat Trung Doan et al. "Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders". In: *NeuroImage: Clinical* 15 (Jan. 1, 2017). ISSN: 2213-1582. DOI: 10.1016/j.nicl.2017.06.014.