medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Developing and sharing polygenic risk scores for 4,206 brain imaging-derived phenotypes for 400,000 UK Biobank subjects not participating in the imaging study

Running title: Imaging PRS in UK Biobank

Xiaochen Yang[1], Bingxuan Li[2], Zirui Fan[3], Dezheng Ding[4], Juan Shu[1], Marylyn D. Ritchie[5,6], Gideon Nave[7], Michael L. Platt[7-9], Tengfei Li[10,11], Hongtu Zhu[12-15]*, and Bingxin Zhao[1,3]*

[1]Department of Statistics, Purdue University, West Lafayette, IN 47907, USA. [2]Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA. [3]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104, USA. [4]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA. [7]Marketing Department, University of Pennsylvania, Philadelphia, PA 19104, USA. [8]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104, USA. [9]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA. [10]Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [11]Biomedical Research Imaging Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [12]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [13]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [14]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [15]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

**Corresponding authors*:

Hongtu Zhu

3105C McGavran-Greenberg Hall, 135 Dauer Drive, Chapel Hill, NC 27599.

E-mail address: htzhu@email.unc.edu Phone: (919) 966-7250

Bingxin Zhao

413 Academic Research Building, 265 South 37th Street, Philadelphia, PA 19104.

E-mail address: bxzhao@upenn.edu Phone: (215) 898-8222

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

The UK Biobank's brain imaging data is an essential resource for clinical research, but its cost and difficulty in obtaining limit the imaging study to only 100,000 participants, leaving the majority of UKB subjects without imaging data. However, because imaging-derived phenotypes (IDPs) are heritable, and most UKB subjects have genetic information available, it's possible to predict IDPs for UKB subjects outside the imaging study using genetic data. To this end, this study systematically developed and evaluated biobank-scale genetic polygenic risk scores (PRS) for 4,206 IDPs from multiple brain imaging modalities and processing pipelines. The results indicate that the majority of IDPs (64.76%, 2,774/4,206) were significantly predicted by PRS developed by subjects with both genetic and imaging data. Moreover, genetically predicted IDPs showed associations with a wide range of complex traits and diseases, with the patterns being consistent across different imaging pipelines. These findings suggest that genetic prediction through PRS is a cost-effective and practical way to make the UKB imaging study more beneficial to a broader population. The PRS data resources developed in this study have been made publicly available through Zenodo and will be returned to the UK Biobank.

Keywords: Brain imaging; Imaging-derived phenotypes; Polygenic risk scores; MRI; UK Biobank

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The large-scale brain imaging data from the UK Biobank (UKB) imaging study has proven to be an immensely valuable resource for characterizing brain structural and functional organizations[1,2]. This data has been instrumental in establishing links with clinical biomarkers[3,4], predicting brain aging[5-7], and facilitating early disease detection[8]. Launched in 2014, the UKB imaging study reached a milestone in 2022 by scanning the multimodal brain magnetic resonance imaging (MRI) of its 50,000th participant. Although it is the world's largest imaging study, the UKB imaging study will ultimately include 100,000 participants[9], leaving 80% (400,000) of the half a million UKB subjects without imaging data. Given the cost and difficulty of collecting additional imaging data, it is crucial to develop strategies that extend the utility of the UKB imaging study to more participants in UKB and a wider population.

The polygenic risk scores (PRS) can be used to predict traits or disease risk for individuals by aggregating genetic information across the genome[10,11]. The development of numerous prediction methods[12], reporting standards[13], genetic data resources[14], and data sharing platforms[15] has enabled the application of PRS to a wide variety of complex diseases and heritable traits. Both family and population-based studies have shown that variation in brain structure and function, as measured by brain MRI, are heritable[2,16-18]. Recent genome-wide association studies (GWAS)[2,19-25] have identified many genetic loci associated with brain imaging-derived phenotypes (IDPs). Consequently, PRS methods can be employed to predict brain IDPs for UKB subjects who are not part of the imaging study. A number of GWAS have investigated and reported the prediction accuracy (out of sample R-squared) of PRS for brain IDPs[19,26,27] in small-scale independent testing data, indicating that genetic data could partially recover variations in brain IDPs, especially when the PRS was developed and applied to the same population or research cohort. These pilot studies demonstrate that genetically-predicted IDPs can serve as valuable proxy imaging biomarkers in the absence of readily available brain MRI data.

In this study, we systematically developed and evaluated PRS for brain IDPs in UKB subjects without imaging data. We examined 4,206 brain IDPs from various imaging modalities and independent processing pipelines, including 3,905 traits generated by the UKB brain imaging team[1,2,20] (UKB Data Category 100, referred to as UKB-Oxford data

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

hereafter, https://open.win.ox.ac.uk/ukbiobank/big40/) and 301 traits processed by BIG-KP[19,21,23] (https://bigkp.org/). These imaging biomarkers spanned major MRI categories, such as structural MRI (sMRI, including regional brain volumes, cortical thicknesses, and surface areas), diffusion MRI (dMRI, with diffusion tensor imaging (DTI) parameters), resting-state functional MRI (rfMRI, featuring amplitude28 and functional connectivity traits), task-based functional MRI (tfMRI, using activation z-statistics), and susceptibility weighted brain MRI (incorporating regional median T2star). The Methods section and **Table S1** provide more detailed information. **Figure 1** presents an overview of the study design. The list of genetic variants and their weights for constructing PRS for brain IDPs can be found at https://github.com/xcyang17/IPRS_UKB.

RESULTS

Developing biobank-scale PRS for 4,206 brain IDPs

To develop and assess the PRS for brain IDPs in UKB subjects without imaging data, we employed data from individuals of white British ancestry with brain IDP information from UKB phases 1 to 3 data releases[2,20,28] for training (average n = 34,224, released up through 2020). We generated GWAS summary statistics of brain IDPs, which were then used as input for the PRS models to create genetically predicted IDPs for all UKB subjects without imaging data (average n = 454,318). An independent hold-out dataset containing brain IDP data served as a test set to evaluate the predictive performance of the generated IDPs (average n = 3,438). We used PRS-CS[29] to construct the PRS, incorporating genotyping data from the UKB study, and 461,488 genetic variants were included in the prediction model after standard genetic data quality controls. The details can be found in **Methods**.

In summary, we discovered that 64.76% (2,774/4,206) of brain IDPs could be significantly predicted after controlling the false discovery rate (FDR) at 5% using the Benjamini-Hochberg procedure. Significant brain IDPs were present in both UKB-Oxford and BIG-KP across all brain MRI modalities (**Fig. 2A** and **Table S2**). Brain IDPs from the same imaging modality exhibited similar prediction accuracy ranges (**Fig. 2B**). For instance, the sMRI modality in BIG-KP consisted of 101 regional brain volumes generated by advanced normalization tools (ANTs)[19,30], with an average prediction R-squared of 1.13% (s.e.=0.10%). The UKB-Oxford contained 1,437 sMRI IDPs, and most subcategories

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

displayed similar prediction R-squared ranges as the BIG-KP ANTs traits. Additionally, there were variations in prediction accuracy across different modalities. Among the three modalities in BIG-KP, dMRI traits (110 DTI parameters from the ENIGMA-DTI TBSS pipeline[21,31,32]) demonstrated significantly higher prediction accuracy compared to sMRI traits (101 ANTs regional brain volumes) or rfMRI traits (90 rfMRI traits using the Glasser360 atlas[21,31,32], $P < 2.2 \times 10^{-16}$, Wilcoxon rank test). Furthermore, we assessed the consistency of prediction performance across different PRS methods by conducting the same analyses on the 301 BIG-KP traits using DBSLMM[21,31,32]. **Figure 2C** reveals that the prediction accuracy of the two methods was consistent across various traits (Correlation = 0.9278, **Table S2**). These results suggest that brain IDPs can be consistently predicted by different PRS prediction methods.

PRS of brain IDPs were widely associated with complex traits

Utilizing the PRS developed for brain IDPs, we conducted association analyses with 265 phenotypes (**Table S3**) on UKB participants who initially lacked brain imaging data (**Methods**). Prior studies have identified associations between various complex traits and diseases using brain imaging data, such as intelligence, blood pressure, and education[3,4]. Our objective is to determine whether imaging-trait relationships can also be uncovered using the PRS of IDPs in the UKB non-imaging cohort. We outline the results for the 301 IDPs from the BIG-KP below.

At the Bonferroni significance level (265 × 301 tests), we discovered 2,053 significant pairs between 97 complex traits and 258 PRS of brain IDPs ($|\beta| > 0.0053$, P range = (1.75×10^{-115}, 6.27×10^{-7})). Out of the 2,053 pairs discovered, 1,922 (93.62%) were replicated in an independent hold-out dataset (**Fig. S1** and **Table S4**). **Figure 3A** shows the pattern of significant IDP-trait connections across different phenotype groups and imaging modalities. The PRS of dMRI traits exhibited the highest percentage of associations, followed by those of sMRI traits and rfMRI traits. Associations were observed across a wide range of phenotypes, including blood biochemistry biomarkers, curated disease phenotypes[33], spirometry, body composition by impedance, and mental health. We provide several examples below, all of which have been replicated.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Significant associations were discovered and replicated between PRS and various disease-related phenotypes, including curated disease phenotypes, family health history, and health and medical history. Among the curated diseases, hyperthyroidism and hypothyroidism were found widely associated with all three imaging modalities ($|\beta| > 0.0104$, $P < 6.02 \times 10^{-7}$). These results were consistent with recent studies[34,35] that reported significant changes in white matter radial diffusivity and axial diffusivity in adult patients with hyperthyroidism/hypothyroidism, which were known to be associated with memory dysfunction. In addition, hypertension and hypercholesterolemia were correlated with PRS of white matter structural connectivity traits and resting functional connectivity traits ($|\beta| > 0.0102$, $P < 5.30 \times 10^{-7}$). Hypertension was mostly correlated with PRS of DTI parameters involving the external capsule and the anterior limb of internal capsule tracts, and there was no association between hypertension and regional brain volumes, which was consistent with a previous study using the UKB brain IDPs[36]. Hypertension can lead to vascular stiffness and impaired cerebral perfusion, which in turn can cause microstructural white matter disruption and stroke[37]. There were significant associations between diabetes and PRS of all three modalities ($|\beta| > 0.0105$, $P < 5.92 \times 10^{-7}$), especially for the DTI parameters of the superior and inferior longitudinal fasciculus tracts and the inferior fronto-occipital fasciculus tract. These findings were consistent with two recent studies that investigated the effects of type 2 diabetes (T2D) on brain white matter[38,39].

We found significant associations with multiple brain-related disorders and the family history of stroke and Alzheimer's disease ($|\beta| > 0.0109$, $P < 2.92 \times 10^{-7}$). Family history of Alzheimer's disease was significantly associated with PRS of DTI parameters of the hippocampal cingulum tracts, consistent with previous findings about changes in the DTI parameters of Alzheimer's disease patients[40-42]. Multiple sclerosis was correlated with all three imaging modalities, such as PRS of DTI parameters of the cingulum and fornix-stria terminalis tracts. Previous research has linked structural damage in the cingulum with subjective fatigue perception in multiple sclerosis[43], and the fornix has been found to be correlated with cognitive impairment in multiple sclerosis patients[44]. We also uncovered widespread associations with brain-related complex traits, including mental health, alcohol use, smoking, cognitive functions, and education. All mental health traits were

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

associated with the PRS of regional brain volumes, and several of them (nervous feelings, visits to doctors/psychiatrists, and neuroticism score) were also widely associated with multiple DTI parameters ($|\beta| > 0.0104$, $P < 6.18 \times 10^{-7}$). Cognitive functions exhibited significant correlations with both DTI parameters and regional brain volumes ($|\beta| > 0.0101$, $P < 5.93 \times 10^{-7}$). For instance, fluid intelligence was positively linked with fractional anisotropy of the uncinate fasciculus, whereas higher fractional anisotropy can improve interhemispheric transfer time, boost information processing speed, and lead to more efficient cognitive functioning and faster reaction[45]. In summary, PRS for brain IDPs provide the opportunity to identify biologically relevant connections between the brain and complex traits and diseases.

Comparison of BIG-KP and UKB-Oxford IDPs in associations with phenotypes

Using the PRS of 3,905 IDPs from the UKB-Oxford database, we repeated association analyses with the 265 phenotypes (**Methods**). Our results confirmed the consistency in PRS-phenotype associations produced by brain IDPs from BIG-KP and UKB-Oxford pipelines. We also discovered new associations from imaging modalities exclusive to the UKB-Oxford. Below we compared the results of UKB-Oxford with those of BIG-KP and highlighted some interesting new associations between PRS and phenotypes.

We found 14,541 significant pairs between 100 phenotypes and 2,814 PRS at the Bonferroni significance level (265 × 3,905 tests; $|\beta| > 0.0056$, P range = (1.01×10^{-135}, 4.83×10^{-8})), 13,899 (95.58%) of which were replicated in an independent dataset (**Figs. 3B** and **S2**, and **Table S5**). Comparing **Figures 3A** and **3B**, both BIG-KP and UKB-Oxford PRS had the most associations in blood biochemistry, curated disease phenotypes, and mental health. The 1,439 sMRI traits in the UKB-Oxford consisted of multiple subcategories, including regional volumes, cortical areas, cortical grey-white contrast, cortical thickness, regional and tissue intensity, regional T2*, and white matter hyperintensity volume (**Table S1**). A high correlation (0.9506) was found between the number of significant associations obtained from the BIG-KP ANTs traits and the UKB-Oxford regional volumes, suggesting that PRS of volumetric measures from the two different pipelines resulted in consistent patterns of phenotypic associations. The other subcategories of sMRI revealed additional associations that were not detected by regional volumes. For example, playing computer

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

games, a possibly addictive behavior, was found negatively associated with the area of the left inferior temporal (β = -0.0115, $P = 4.35 \times 10^{-8}$). It was reported that young male adults playing Internet video games had smaller inferior temporal gyri[46]. The associations detected by the 675 UKB-Oxford dMRI traits (tract-skeleton and probabilistic tractography traits) highly overlapped with those of the BIG-KP DTI parameters, with the correlation between the number of significant associations being 0.9797. The PRS of the dMRI traits had additional significant associations with rheumatoid arthritis and liver biomarkers (such as gamma-glutamyl transferase and direct bilirubin) in multiple white matter tracts ($|\beta| > 0.0117$, $P < 4.30 \times 10^{-8}$). Both gamma-glutamyl transferase and direct bilirubin were related to rheumatoid arthritis[47,48], and previous studies have shown brain atrophy in rheumatoid arthritis patients[49].

A wide range of phenotypes was associated with PRS of 1,777 rfMRI IDPs, including the family history of Alzheimer's disease, neuroticism, cardiovascular problems, and blood biomarkers ($|\beta| > 0.0112$, $P < 4.22 \times 10^{-8}$). For example, the family history of Alzheimer's disease was associated with the PRS in the visual network and the three core cognitive networks (the central executive, default mode, and salience networks) ($\beta < -0.0115$, $P < 4.67 \times 10^{-8}$). Previous studies found that Alzheimer's disease progressively reduced visual functional network connectivity[50], and MRI of the three core cognitive networks are known to be predictive of Alzheimer's disease[51-53]. We also detected multiple associations between neuroticism and the PRS of rfMRI IDPs in the cerebellum ($|\beta| > -0.0131$, $P < 2.08 \times 10^{-8}$). Cerebellum plays an important role in motion control and is involved in cognitive functions, and previous studies showed functional connectivity of the cerebellum was highly involved in neuroticism[54]. Overall, the BIG-KP and UKB-Oxford IDPs provide consistent association patterns across different categories of phenotypes.

Concordance between brain IDPs and their PRS

In this section, we conducted an analysis of phenotypic associations between IDPs and phenotypes on UKB subjects with brain imaging data (average n = 34,870, **Methods**). We then compared the IDP-phenotype associations in the UKB imaging cohort with the PRS-phenotype associations in the UKB non-imaging cohort. At the FDR 5% level (265 × 301 tests), 4,717 pairs between 206 phenotypes and 297 IDPs were discovered and replicated

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

($|\beta| > 0.0076$, P range = (1.18×10^{-123}, 8.19×10^{-3}); **Figs. 4A** and **S3**, and **Table S6**). Out of these 4,717 IDP-phenotype associations, 1,383 pairs between 121 phenotypes and 266 PRS were significant at the FDR 5% level (**Figs. 4B** and **S4**, and **Table S7**). That is, PRS associations recovered 29.32% (1,383/4,717) of the IDP associations, corresponding to 58.74% (121/206) of the phenotypes and 89.56% (266/297) of the imaging traits. The distribution of the IDP signals in **Figure 4A** was in more diverse phenotype groups than that of the PRS signals in **Figure 4B**. Both IDP and PRS results were most abundant in blood biochemistry, curated disease phenotypes, and mental health traits, and the distribution of signals in each imaging modality was consistent. Among the 1,383 pairs that were significant in both IDP and PRS analyses, the correlation between their regression coefficients was 0.5685, and 78.16% (1,081/1,383) had regression coefficients in the same direction. The correlation among regression coefficients reduced to 0.4389 among all the 4,717 IDP-phenotype associations. These results suggest that the majority of PRS associations have the same signs as the IDP associations and their regression coefficients are partially overlapped.

Significant associations with various brain disorders, including stroke, multiple sclerosis, depression, and migraine, were identified by both brain IDPs ($|\beta| > 0.0168$, $P < 4.62 \times 10^{-3}$) and their PRS ($|\beta| > 0.0064$, $P < 2.48 \times 10^{-3}$). In multiple white matter tracts, significant positive associations with stroke were found with mean diffusivity and residual diffusivity and their PRS (IDP $\beta > 0.0327$, $P < 1.21 \times 10^{-7}$; PRS $\beta > 0.0064$, $P < 2.48 \times 10^{-3}$). These findings are consistent with the known impairment of white matter and motor deficits following stroke[55]. Similar to a recent study[56], the mean diffusivity of the superior fronto-occipital fasciculus and its PRS had positive associations with depression (IDP $\beta = 0.0390$, $P = 1.24 \times 10^{-10}$; PRS $\beta = 0.0079$, $P = 1.89 \times 10^{-4}$). The brain-related complex traits that were associated with both brain IDPs and PRS included most mental health traits, cognitive functions, and electronic device usage, such as time spent watching TV, weekly use of mobile phone, and length of mobile phone use (IDP $|\beta| > 0.0174$, $P < 4.52 \times 10^{-3}$; PRS $|\beta| > 0.0059$, $P < 4.83 \times 10^{-3}$). In addition, we found that some brain orders were only significantly associated with brain IDPs, and not with PRS, such as bipolar disorder, Parkinson's disease, and epilepsy ($|\beta| > 0.0015$, $P < 7.68 \times 10^{-3}$). For example, there were strong negative associations between bipolar disorder and the mean fractional anisotropy

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of the body and genus of corpus callosum (β < -0.0197, P < 5.82 × 10^{-4}), which was consistent with findings reported in other studies[57,58]. In summary, we explored the overlaps between IDP and PRS phenotypic associations and confirmed that PRS can partially recover the imaging associations with brain-related diseases and complex traits. The PRS can be used as proxy imaging biomarkers when brain MRI data are unavailable.

DISCUSSION

In this study, we generated PRS for 4,206 brain IDPs for UKB subjects without imaging data. These PRS have been investigated in relation to a wide range of phenotypes and it was confirmed that they can provide biologically relevant information to brain-related complex traits and diseases. We found consistent predictive accuracy and association patterns across IDPs from different pipelines, such as the volumetric measures in UKB-Oxford and BIG-KP. The PRS of brain IDPs partially recovered previously known associations generated from imaging data. It is possible to detect almost 30% of the IDP associations using their PRS proxy data, and the majority of these PRS associations have the same sign as the IDP associations. We have provided the data resources so that users can easily reconstruct PRS in the UKB database.

When real brain imaging data are not available, the PRS can be used as genetically predicted variables for brain structure and function. However, as shown in our prediction and association analyses, the PRS is only able to partially reconstruct the imaging phenotypes. It has generally been observed that PRS has demonstrated imperfect performance in predicting the most complex traits and diseases, which can be attributed to a number of factors, including a limited number of training samples, heritability, and weak genetic effects[59]. Another challenge in PRS applications lies in ancestry and population differences. As the current UKB imaging cohort had the majority of the subjects of European ancestry, generating PRS in non-UKB and/or non-European studies may have further reduced performance[60]. More powerful PRS methods that better account for PRS limitations and cohort differences may result in more informative PRS for potential clinical applications.

METHODS

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods are available in the ***Methods*** section.

Note: One supplementary information pdf file and one supplementary table zip file are available.

ACKNOWLEDGEMENTS

The study has been partially supported by funding from the Wharton Dean's Research Fund and Analytics at Wharton, as well as start-up funds from Purdue Statistics Department. This research has been conducted using the UK Biobank resource (application number 76139), subject to a data transfer agreement. We would like to thank the individuals who represented themselves in the UK Biobank for their participation and the research teams for their efforts in collecting, processing, and disseminating these datasets. We would like to thank the research computing groups at the University of North Carolina at Chapel Hill, Purdue University, and the Wharton School of the University of Pennsylvania for providing computational resources and support that have contributed to these research results.

REFERENCES

1. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400-424 (2018).
2. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
3. Miller, K.L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience* **19**, 1523-1536 (2016).
4. Smith, S.M. & Nichols, T.E. Statistical challenges in "big data" human neuroimaging. *Neuron* **97**, 263-268 (2018).
5. Cole, J.H. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. *Neurobiology of aging* **92**, 34-42 (2020).
6. Cox, S.R. *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nature communications* **7**, 13629 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

7. 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).
8. Azevedo, T. *et al.* Identifying healthy individuals with Alzheimer neuroimaging phenotypes in the UK Biobank. *medRxiv* (2022).
9. Littlejohns, T.J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature communications* **11**, 1-12 (2020).
10. Lewis, C.M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine* **12**, 1-11 (2020).
11. Torkamani, A., Wineinger, N.E. & Topol, E.J. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* **19**, 581-590 (2018).
12. Ma, Y. & Zhou, X. Genetic prediction of complex traits with polygenic scores: a statistical review. *Trends in Genetics* **37**, 995-1011 (2021).
13. Wand, H. *et al.* Improving reporting standards for polygenic scores in risk prediction studies. *Nature* **591**, 211-219 (2021).
14. Uffelmann, E. *et al.* Genome-wide association studies. *Nature Reviews Methods Primers* **1**, 1-21 (2021).
15. Lambert, S.A. *et al.* The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nature Genetics* **53**, 420-425 (2021).
16. Zhao, B. *et al.* Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cerebral Cortex* **29**, 2904-2914 (2018).
17. Kochunov, P. *et al.* Homogenizing estimates of heritability among SOLAR-Eclipse, OpenMx, APACE, and FPHI software packages in neuroimaging data. *Frontiers in Neuroinformatics* **13**, 16 (2019).
18. Thompson, P.M. *et al.* 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**, 1-28 (2020).
19. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nature genetics* **51**, 1637-1644 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

20. Smith, S.M. *et al.* An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature neuroscience* **24**, 737-745 (2021).
21. Zhao, B. *et al.* Common genetic variation influencing human white matter microstructure. *Science* **372**(2021).
22. Grasby, K.L. *et al.* The genetic architecture of the human cerebral cortex. *Science* **367**(2020).
23. Zhao, B. *et al.* Genetic influences on the intrinsic and extrinsic functional organizations of the cerebral cortex. *medRxiv* (2021).
24. Hofer, E. *et al.* Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nature communications* **11**, 1-16 (2020).
25. Satizabal, C.L. *et al.* Genetic architecture of subcortical brain structures in 38,851 individuals. *Nature genetics* **51**, 1624-1636 (2019).
26. Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Molecular Psychiatry* (2019).
27. Zhao, B. *et al.* Common variants contribute to intrinsic human brain functional networks. *Nature Genetics* **in press** (2022).
28. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
29. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. & Smoller, J.W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* **10**, 1776 (2019).
30. Avants, B.B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033-2044 (2011).
31. Jahanshad, N. *et al.* Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA–DTI working group. *Neuroimage* **81**, 455-469 (2013).
32. Kochunov, P. *et al.* Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and megaanalytical approaches for data pooling. *Neuroimage* **95**, 136-150 (2014).

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

33. Dey, R. *et al.* An efficient and accurate frailty model approach for genome-wide survival association analysis controlling for population structure and relatedness in large-scale biobanks. *bioRxiv* (2020).

34. Aslan, K. *et al.* Diffusion tensor imaging in hyperthyroidism: assessment of microstructural white matter abnormality with a tract-based spatial statistical analysis. *Acta Radiol* **61**, 1677-1683 (2020).

35. Singh, S. *et al.* Diffusion tensor tractography in hypothyroidism and its correlation with memory function. *J Neuroendocrinol* **26**, 825-33 (2014).

36. Suzuki, H. *et al.* Abnormal brain white matter microstructure is associated with both pre-hypertension and hypertension. *PLoS One* **12**, e0187600 (2017).

37. Karlsgodt, K.H. *et al.* Diffusion tensor imaging of the superior longitudinal fasciculus and working memory in recent-onset schizophrenia. *Biol Psychiatry* **63**, 512-8 (2008).

38. Huang, L. *et al.* Abnormalities of Brain White Matter in Type 2 Diabetes Mellitus: A Meta-Analysis of Diffusion Tensor Imaging. *Front Aging Neurosci* **13**, 693890 (2021).

39. Ma, T. *et al.* Gray and white matter abnormality in patients with T2DM-related cognitive dysfunction: a systemic review and meta-analysis. *Nutr Diabetes* **12**, 39 (2022).

40. Bozzali, M. *et al.* Damage to the cingulum contributes to Alzheimer's disease pathophysiology by deafferentation mechanism. *Hum Brain Mapp* **33**, 1295-308 (2012).

41. Rao, Y.L. *et al.* Hippocampus and its involvement in Alzheimer's disease: a review. *3 Biotech* **12**, 55 (2022).

42. Mayo, C.D. *et al.* Relationship between DTI metrics and cognitive function in Alzheimer’s disease. *Frontiers in aging neuroscience* **10**, 436 (2019).

43. Pardini, M. *et al.* Cingulum bundle alterations underlie subjective fatigue in multiple sclerosis. *Mult Scler* **21**, 442-7 (2015).

44. Keser, Z. *et al.* Quantitative Limbic System Mapping of Main Cognitive Domains in Multiple Sclerosis. *Front Neurol* **9**, 132 (2018).

45. Penke, L. *et al.* Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular psychiatry* **17**, 1026-1030 (2012).

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

46. Han, D.H. *et al.* Brain activity and desire for Internet video game play. *Compr Psychiatry* **52**, 88-95 (2011).

47. Spooner, R.J., Smith, D.H., Bedford, D. & Beck, P.R. Serum gamma-glutamyltransferase and alkaline phosphatase in rheumatoid arthritis. *J Clin Pathol* **35**, 638-41 (1982).

48. Peng, Y.F., Wang, J.L. & Pan, G.G. The correlation of serum bilirubin levels with disease activity in patients with rheumatoid arthritis. *Clin Chim Acta* **469**, 187-190 (2017).

49. Schrepf, A. *et al.* A multi-modal MRI study of the central response to inflammation in rheumatoid arthritis. *Nat Commun* **9**, 2243 (2018).

50. Huang, J., Beach, P., Bozoki, A. & Zhu, D.C. Alzheimer's Disease Progressively Reduces Visual Functional Network Connectivity. *J Alzheimers Dis Rep* **5**, 549-562 (2021).

51. Daigle, K.M., Pietrzykowski, M.O., Waters, A.B., Swenson, L.P. & Gansler, D.A. Central Executive Network and Executive Function in Patients With Alzheimer's Disease and Healthy Individuals: Meta-Analysis of Structural and Functional MRI. *J Neuropsychiatry Clin Neurosci* **34**, 204-213 (2022).

52. Eyler, L.T. *et al.* Resting State Abnormalities of the Default Mode Network in Mild Cognitive Impairment: A Systematic Review and Meta-Analysis. *J Alzheimers Dis* **70**, 107-120 (2019).

53. Schultz, A.P. *et al.* Longitudinal degradation of the default/salience network axis in symptomatic individuals with elevated amyloid burden. *Neuroimage Clin* **26**, 102052 (2020).

54. Hsu, W.T., Rosenberg, M.D., Scheinost, D., Constable, R.T. & Chun, M.M. Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Soc Cogn Affect Neurosci* **13**, 224-232 (2018).

55. Puentes, S. *et al.* Internal capsule stroke in the common marmoset. *Neuroscience* **284**, 400-411 (2015).

56. van Velzen, L.S. *et al.* White matter disturbances in major depressive disorder: a coordinated analysis across 20 international cohorts in the ENIGMA MDD working group. *Mol Psychiatry* **25**, 1511-1525 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

57. Favre, P. *et al.* Widespread white matter microstructural abnormalities in bipolar disorder: evidence from mega- and meta-analyses across 3033 individuals. *Neuropsychopharmacology* **44**, 2285-2293 (2019).

58. Barysheva, M., Jahanshad, N., Foland-Ross, L., Altshuler, L.L. & Thompson, P.M. White matter microstructural abnormalities in bipolar disorder: A whole brain diffusion tensor imaging study. *Neuroimage Clin* **2**, 558-68 (2013).

59. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS genetics* **9**, e1003348 (2013).

60. Wang, Y. *et al.* Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature communications* **11**, 1-9 (2020).

61. 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).

62. Glasser, M.F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171-178 (2016).

63. Ji, J.L. *et al.* Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* **185**, 35-57 (2019).

64. Beckmann, C.F. & Smith, S.M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging* **23**, 137-152 (2004).

65. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks* **10**, 626-634 (1999).

66. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* **51**, 1749-1755 (2019).

67. Yang, S. & Zhou, X. Accurate and Scalable Construction of Polygenic Scores in Large Biobank Data Sets. *Am J Hum Genet* **106**, 679-693 (2020).

METHODS

Imaging traits.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The data used in our study was obtained from the UK Biobank (UKB) study, which recruited around 500,000 individuals between the ages of 40 and 69 between 2006 and 2010[61] (https://www.ukbiobank.ac.uk/). The ethics approval of the UKB study was obtained from the North West Multicentre Research Ethics Committee (approval number: 11/NW/0382). We used a total of 4,206 brain imaging-derived phenotypes (IDPs) from the UKB study, which consisted of 301 BIG-KP[19,21,23] (https://bigkp.org/) and 3,905 UKB-Oxford[1,2,20] (https://open.win.ox.ac.uk/ukbiobank/big40/) traits. BIG-KP traits were divided into three groups. First, we obtained 101 regional brain volumes[19] from structural MRI (sMRI) images by applying the advanced normalization tools[30] (ANTs). Second, we generated 110 tract-averaged diffusion tensor imaging (DTI) parameters from diffusion MRI (dMRI) using the ENIGMA-DTI pipeline[31,32]. Third, for resting-state fMRI (rsfMRI), we partitioned the cerebral cortex into 360 brain areas using the Glasser360 atlas[62]. We obtained 90 functional activity (amplitude) and functional connectivity (edge) traits for 12 functional networks[63]. The UKB-Oxford had 1,437 IDPs from sMRI, 675 from dMRI, 1,777 from rsfMRI, and 16 from task-based functional MRI (tfMRI). The sMRI IDPs consisted of FIRST (Category 1102), FAST (Category 1101), FreeSurfer ASEG (Category 190), FreeSurfer BA exvivo (Category 195), FreeSurfer a2009s (Category 197), FreeSurfer DKT (Category 196), FreeSurfer desikan gw (Category 194), FreeSurfer desikan pial (Category 193), FreeSurfer desikan white (Category 192), FreeSurfer subsegmentation (Category 191), regional T2* (Category 109), and white matter hyperintensity volume (Category 112). The 675 dMRI IDPs included 432 from Category 134 and 243 from Category 135. The 1,777 rsfMRI IDPs included 76 amplitude (node) traits and 1,701 functional connectivity (edge) traits from whole brain spatial independent component analysis[1,64,65] (Category 111). Lastly, there were 16 tsfMRI IDPs from Category 106. The image acquisition, preprocessing procedures, and quality controls were detailed in the UKB Brain Imaging Documentation (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf). See **Table S1** for the complete ID list of all brain IDPs.

PRS constructions.

We performed the following genetic quality controls for the set of subjects with both brain IDPs and genetic data[23]: 1) removed individuals with missing genotype rate > 0.1; 2) removed variants with missing genotype rate > 0.1; 3) removed variants with minor allele

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

frequency (MAF) < 0.01; and 4) removed variants that failed the Hardy-Weinberg test for equilibrium at 1×10^{-7} level. Using individuals of white British ancestry, the GWAS was performed using linear mixed effect models via fastGWA[66] (average *n* = 34,224). The adjusting covariates included age (at imaging), age-squared, sex, the interaction between age and sex, the interaction between age-squared and sex, first 40 genetic principal components[28] (PCs), estimated total intracranial volume (eTIV), head motion measurements and their squares, brain position measurements and their squares, and volumetric scaling. Additionally, for regional brain volume IDPs, the total brain volume (TBV) was included as an adjusting covariate. For TBV, the eTIV and volumetric scaling were not included as covariates. With the GWAS summary statistics as input, we applied PRS-CS[29] and DBSLMM[67] to obtain the effect sizes. The hyperparameters of both methods were the default values and/or the automatically tuned values. We then used PLINK to generate risk scores in testing data by summarizing across genetic variants, weighed by their effect sizes estimated from PRS-CS[29] and DBSLMM[67].

The prediction accuracy of PRS was measured by the incremental R-squared, which was the additional phenotypic variation that can be explained by the PRS while adjusting for the effects of covariates in a linear regression model. The covariates included age, age-squared, sex, the interaction between age and sex, the interaction between age-squared and sex, and the first 40 genetic PCs. The prediction accuracy was estimated in a dataset consisting of unrelated UKB individuals of non-British ancestry with brain IDP data (average *n* = 3,200).

PRS-phenotype and IDP- phenotype association analyses.

We employed a discovery-replication approach to examine associations between PRS and phenotypes in UKB participants without brain IDPs. We randomly selected 70% of UKB British white individuals (average *n* = 202,893) as the discovery dataset for PRS-phenotype associations, while the remaining 30% of UKB British white individuals, all UKB white but non-British individuals, and all non-white individuals (average *n* = 129,333) were used as the replication dataset. We treated the values greater than five times the median absolute deviation from the median as outliers and removed these values. A total of 265 UKB phenotypes were tested, which represented a wide range of traits from various trait

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

domains. Specifically, the 265 UKB phenotypes included 24 mental health traits (Category 100060), 5 cognitive traits (Category 100026), 12 physical activity traits (Category 100054), 6 electronic device use traits (Category 100053), 8 sun exposure traits (Category 100055), 3 sexual factor traits (Category 100056), 3 social support traits (Category 100061), 12 family history of diseases (Category 100034), 21 diet traits (Category 100052), 9 alcohol drinking traits (Category 100051), 6 smoking traits (Category 100058), 34 blood biochemistry biomarkers (Category 17518), 3 blood pressure traits (Category 100011), 3 spirometry traits (Category 100020), 32 early life factors (Categories 135, 100033, 100034, and 100072), 9 greenspace and coastal proximity (Category 151), 2 hand grip strength (Category 100019), 13 residential air pollution traits (Category 114), 5 residential noise pollution traits (Category 115), 2 body composition traits by impedance (Category 100009), 4 health and medical history traits (Category 100036), 3 female specific factors (Category 100069), 1 education trait (Category 100063), and 57 curated disease phenotypes based on Dey, et al. [33] (**Table S3**).

Association testing was then conducted to examine the relationship between the 4,206 IDP-derived PRS generated by PRS-CS and the 265 UKB phenotypes. To investigate the PRS-phenotype associations, we conducted a linear regression analysis, adjusting for the same set of covariates separately in the discovery set and the replication set. The adjusted covariates included age, age-squared, sex, the interaction between age and sex, the interaction between age-squared and sex, and 40 genetic PCs. Specifically, we regressed the IDP-derived PRS onto the UKB phenotypes and calculated *P* values using a two-sided t-test. We prioritized the results that met the following three criteria: 1) significant after Bonferroni correction in the discovery dataset, 2) significant at a nominal significance level (0.05) in the replication dataset, and 3) had regression coefficients with matching directions in both the discovery and replication datasets.

We analyzed the associations between the 301 BIG-KP IDPs and the 265 UKB phenotypes in a discovery-replication design. Specifically, the discovery set included all unrelated white British subjects from UKB phases 1 to 3 data releases, which was similar to the training GWAS dataset. The replication set consisted of all the rest of the non-discovery unrelated subjects from UKB phases 1 to 3 data releases and all unrelated subjects from

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

UKB phase 4 data release. We performed linear regression on the discovery dataset and the replication dataset, respectively, by regressing the BIG-KP IDP on the UKB phenotype. We adjusted for the same set of covariates as used in the GWAS analysis. We reported *P* values from the two-sided t-test and prioritized those that met the following three criteria: 1) significant in the discovery dataset after controlling the false discovery rate (FDR) at a 5% level with the Benjamini-Hochberg procedure, 2) significant at a nominal significance level in the replication dataset, and 3) had regression coefficients with matching directions in both the discovery and replication datasets.

Code availability.

We made use of publicly available software and tools. The list of genetic variants and their weights used to construct PRS for brain IDPs are available at https://github.com/xcyang17/IPRS_UKB.

Data availability.

The PRS data resources have been made publicly available at Zenodo (https://doi.org/10.5281/zenodo.7709788). The individual-level data used in this study can be obtained from https://www.ukbiobank.ac.uk/.

Figure legends.

Fig. 1 Study overview.

An overview of the study design. From raw brain imaging data, 4,206 brain imaging-derived phenotypes (IDPs) were extracted using various pipelines in previous UKB-Oxford and BIG-KP projects. Multiple brain magnetic resonance imaging (MRI) modalities were considered, including structural MRI, diffusion MRI, resting-state functional MRI (rfMRI), and task-based functional MRI (tfMRI). PRS of brain IDPs were generated for non-imaging subjects (subjects without brain MRI data) using genetic predictors. Prediction accuracy was examined and associations with a wide range of phenotypes were performed.

Fig. 2 Prediction analysis.

(A) The proportion of brain IDPs significantly predicted by their corresponding PRS (generated by PRS-CS) in each imaging modality, after controlling the false discovery rate

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(FDR) rate at 5% by the Benjamini-Hochberg procedure and at the Bonferroni 0.05 level, respectively. **(B)** The incremental prediction accuracy of PRS (generated by PRS-CS) in each imaging modality. **(C)** Comparison of incremental prediction accuracy of PRS using PRS-CS and DBSLMM.

Fig. 3 PRS-phenotype association analysis.

(A) We visualized the pattern of significant PRS-phenotype associations (after Bonferroni correction for multiple testing) between various groups of UKB phenotypes (upper panel) and three imaging modalities in BIG-KP (lower panel). **(B)** We visualized the pattern of significant PRS-phenotype associations (after Bonferroni correction for multiple testing) between various groups of UKB phenotypes (upper panel) and four imaging modalities in UKB-Oxford (lower panel). Each segment in the upper half of the visualization represents the number of significant signals found within a phenotype group, while each segment in the lower half represents the number of significant signals found within an imaging modality. These segments are color-coded to correspond to either a phenotype group or an imaging modality.

Fig. 4 Phenotypic associations with PRS and IDPs.

(A) We visualized the pattern of significant IDP-phenotype associations (at a false discovery rate of 5% level) between various groups of UKB phenotypes (upper panel) and three imaging modalities in BIG-KP (lower panel). **(B)** We visualized the pattern of significant PRS-phenotype associations (at a false discovery rate of 5% level) between various groups of UKB phenotypes (upper panel) and three imaging modalities in BIG-KP (lower panel). Each segment in the upper half of the visualization represents the number of significant signals found within a phenotype group, while each segment in the lower half represents the number of significant signals found within an imaging modality. These segments are color-coded to correspond to either a phenotype group or an imaging modality.

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

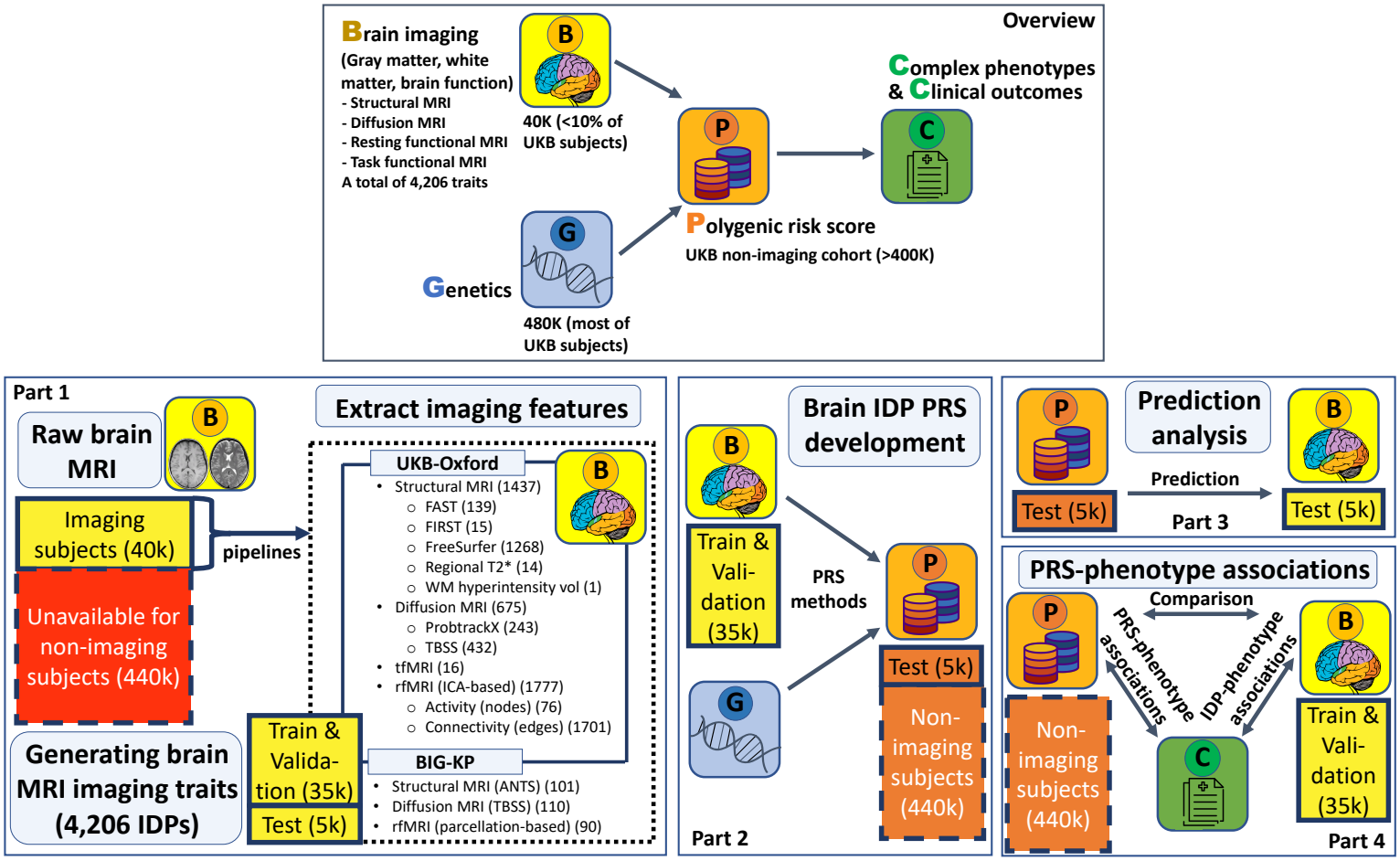

Figure 1

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

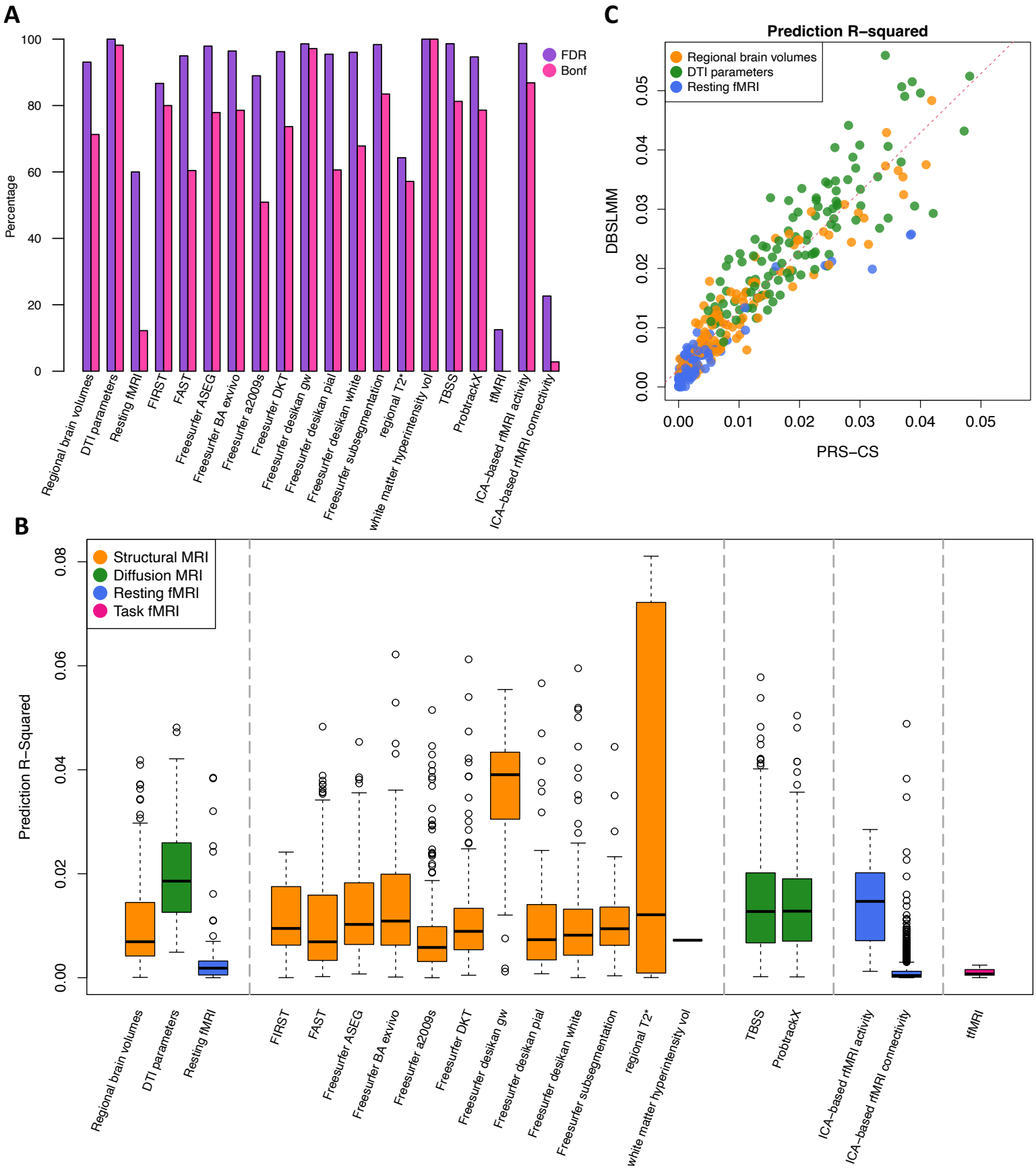

Figure 2

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

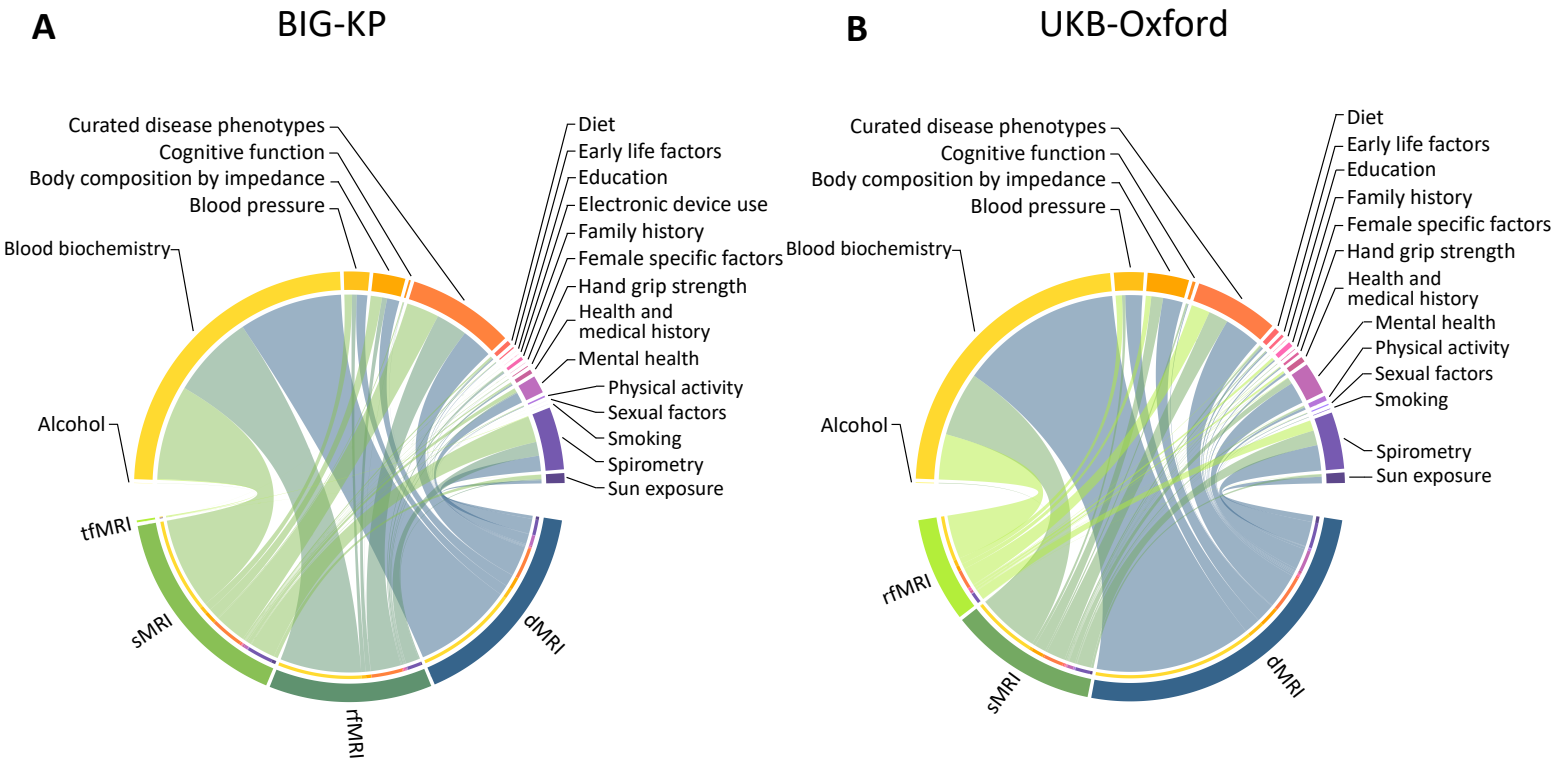

Figure 3

medRxiv preprint doi: https://doi.org/10.1101/2023.04.18.23288769; this version posted April 20, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

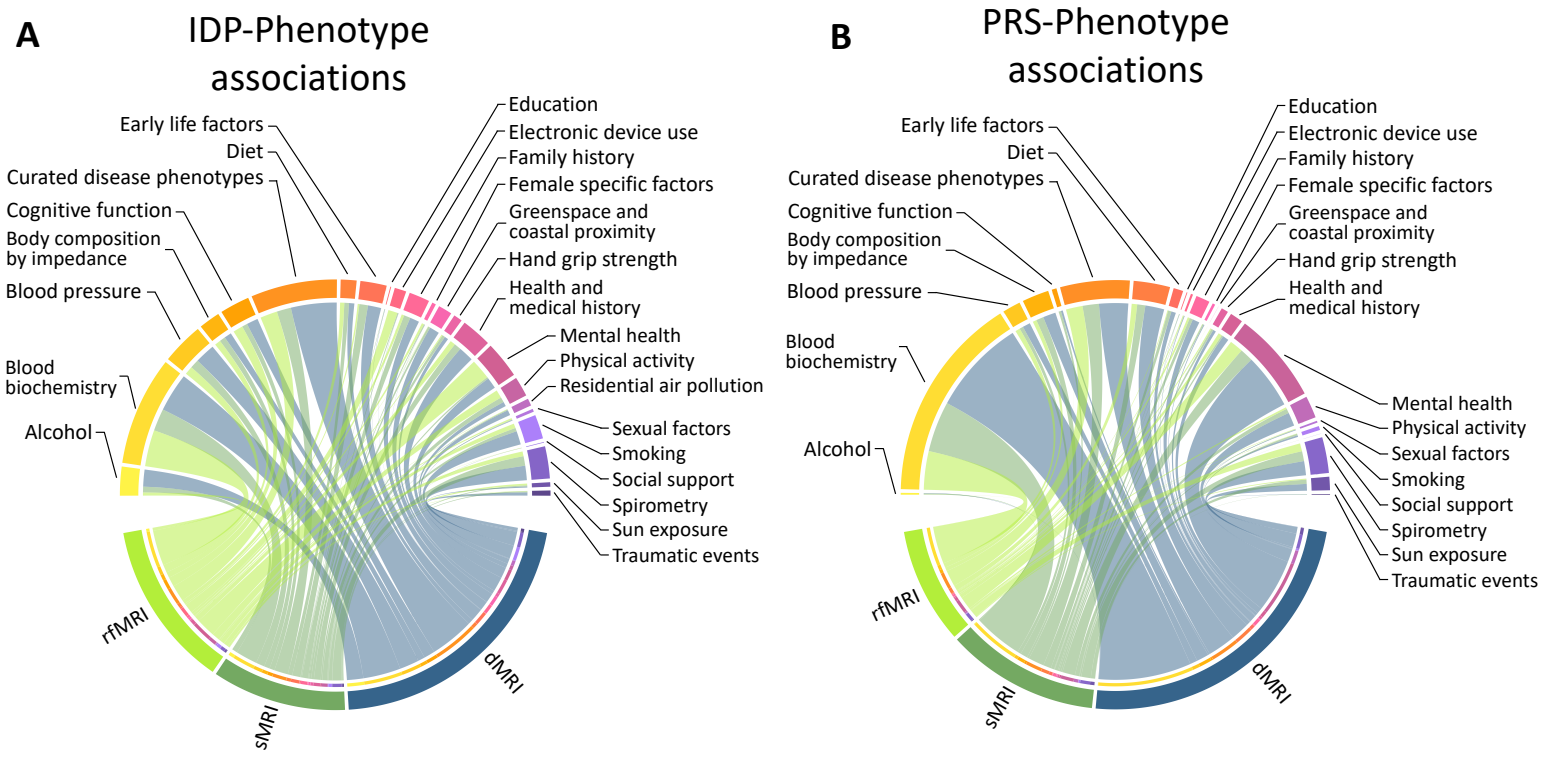

Figure 4