Modelling cognitive outcomes in the UK Biobank: education, noradrenaline and frontoparietal networks

Laura Bravo-Merodio[1,2], Jackie A. Williams[1,2,3], Dominic Russ[1,2], Georgios Gkoutos[1,2], Meadhbh Brosnan[4,5], Mark A. Bellgrove[4], Magdalena Chechlacz[6,7]

[1]Institute of Cancer and Genomic Sciences, Centre for Health Data Science, University of Birmingham, Birmingham, UK; [2]Institute for Translational Medicine, University of Birmingham, Birmingham, UK; [3]Mammalian Genetics Unit, Medical Research Council Harwell Institute, Harwell, UK; [4]Turner Institute for Brain and Mental Health and School of Psychological Sciences, Monash University, Melbourne, Victoria, Australia; [5]School of Psychology,University Collge Dublin, Dublin, Ireland; [6]Centre for Human Brain Health, University of Birmingham, Birmingham, UK; [7]School of Psychology, University of Birmingham, Birmingham, UK

ABSTRACT

Cognitive reserve (CR) measured by proxies of life experiences is thought to account for the heterogeneity in cognitive ageing. One hypothesis is that cognitively enriched environments acting on the noradrenergic system strengthen the right fronto-parietal networks to facilitate cognitive reserve and maintain cognition throughout the lifetime. We used machine learning and mediation analysis to model interactions between neurobiological features (genetic variants in noradrenergic signalling, structural and functional fronto-parietal connectivity) and CR (by proxy of education) on cognitive outcomes (general cognitive ability score) in the UK Biobank cohort. We show that: (1) interactions between CR and neurobiological variables more fully explain cognitive outcomes than either factor alone; (2) among the neurobiological features selected using variable importance testing, measures of fronto-parietal connectivity are the strongest mediators of the effect of CR on cognitive outcomes. Our findings offer novel insights into neurobiological basis of CR by pointing to two between-networks

connectivity measures, representing connections linking the default mode network with the right fronto-parietal network, and with the executive control network, as the key facilitators of CR.

INTRODUCTION

The global demographic shift over the last few decades leading to increased proportion of older adults resulted in steadfastly raising numbers of people at risk of cognitive decline and dementia. This has long-lasting impact not only on the individual quality of life but also major economic impact on the societies worldwide (Livingston et al., 2020). While the ageing process is inevitable and comes with expectation of worsening in cognitive functioning, the literature examining cognitive ageing points to a striking heterogeneity in the rate of age-related deterioration in cognitive decline. It means that some older adults retain high levels of mental capacity, some undergo a gradual drop in cognitive functioning, and finally some experience a sharp cognitive decline (dementia) hindering their ability to undertake basic daily activities and affecting quality of life (Rapp and Amaral 1992, Hayden et al., 2011, Norton et al., 2014; Cabeza et al., 2018; Stern et al., 2020). To keep up with the ageing population and to enable advances in medical/social care a better understanding of factors affecting the differential trajectories of cognitive ageing is urgently needed (Prince et al., 2015; Livingstone et al., 2020).

Cognitive ageing is a highly complex process caused by a convergence of multiple neurobiological and neurophysiological changes influenced by environmental and genetic factors as well as gene-environment interactions (e.g., Harris and Deary, 2011; Davies et al., 2015; Lu et al., 2016; Tucker-Drob et al., 2014, Papenberg et al., 2016). Among environmental influences, sociodemographic and lifestyle factors seem to play a critical role (for review see Stern et al., 2020; Cabeza et al., 2018). While it is well evidenced that the cognitive ageing is a result of structural and functional brain changes causing deterioration of cognitive performance, the interpretation of multifactorial influences on age-related changes in neural networks underlying the differential trajectories of cognitive ageing presents an ongoing

challenge. There is a growing evidence that older adults, who have engaged across lifespan in cognitively and/or socially enriched environments exhibit greater resilience to cognitive decline and maintain better brain health and cognitive function in later year. The concept of cognitive reserve (CR) has been proposed to account for at least some the observed heterogeneity in healthy (non-dementia) cognitive ageing as well as susceptibility to dementia, pathological decline (Stern et al., 2020; Cabeza et al., 2018). Within this framework, CR is measured by proxies of life experiences, in particular education but also occupation and leisure activities (Cabeza et al., 2018; Valenzuela and Sachdev, 2006a,b; Opdebeeck et al., 2015; Nucci et al., 2012).

While the concept of CR offers appealing framework for understanding cumulative neural changes due to lifelong environmental influences, including specifically socio-behavioural factors, and their interplay with genetic variability, yet to date few theoretical and experimental attempts have been made in understanding the neurobiological origins of cognitive reserve in healthy ageing (Cabeza et al., 2018; Robertson et al., 2014, Brosnan et al., 2018, Brosnan et al., 2022; Shalev at al., 2020; Plini et al., 2021). Here, we used one such theoretical framework (Robertson's theory; Robertson et al., 2014) of CR to test machine learning methods for a potential to tease apart combinatorial effects of environmental and genetic influences on brain networks and cognitive outcomes in a large cross-sectional UK Biobank population. As UK Biobank has been created as a large-scale epidemiological resource with extensive sociodemographic questionnaires, physical measures, medical records, neuroimaging, and genetic data from middle and old age participants (Sudlow et al., 2015), it enables to tease apart multifactorial influences on cognitive ageing, using data science approaches on a scale not feasible before.

Robertson's theory of CR proposes that life experiences acting on the noradrenergic system strengthen the right fronto-parietal networks to facilitate cognitive reserve and maintain cognition throughout lifetime (Robertson et al., 2013; Robertson et al., 2014). This theory combines Stern's observations on the effects of education in Alzheimer's patients with animal research. Animal studies provide evidence that across the life-span neural networks are heavily subjected to regulatory influences of neuromodulators (dopamine, noradrenaline, serotonin, and acetylcholine). These neuromodulators are thought to maintain functional dynamics among large-scale brain networks and to optimize cognitive performance by signalling environmental inputs (for review see Avery and Krichmar, 2017). In line with animal research the Robertson's theory stipulates that life experiences trigger continues engagement of several core cognitive process, all of which functionally rely on locus coeruleus—noradrenergic system and right fronto-parietal networks (Robertson et al., 2013; Robertson et al., 2014). A recent study linking cognitive performance and brain health in both neurotypical older adults and Alzheimer's patients to the volumetric estimates of locus coeruleus has provided compelling evidence that indeed the noradrenergic system might underpin the CR (Plini et al., 2021). There is also evidence in support of right lateralised cognitive reserve network (van Loenhoud et al., 2017; Shalev et., 2020, Brosnan et la., 2018; Brosnan et al., 2022). For example, our group has previously shown that life experiences, including education, mitigate age-related cognitive deficits by (i) preserving grey matter withing the right fronto-parietal regions (Shalev et al., 2020) and (ii) offsetting age-related axonal dispersion within the right fronto-parietal white matter (Brosnan et al., 2022). It should be noted that none of the previous studies examining neurobiological basis of Robertson's theory of CR simultaneously tapped into the noradrenergic system and the right fronto-parietal networks.

Growing evidence suggests that genetic variations, leading to either elevated or decreased levels of neuromodulators have impact on the functional dynamics of the neural networks, and underlie inter-individual variability in cognitive abilities across human lifetime (Lindenberger et al., 2008; Papenberg et al., 2015). Thus, along these lines the Robertson's theory of CR presents an interesting framework for understanding mechanism of cumulative environmental and genetic influences on functional neural networks and whether such interplay contributes to CR. It enables to stipulate that genetic variability, enhancing neuromodulation across lifespan, and strengthening the right fronto-parietal networks in response to life experiences could be advantageous in offsetting age-related cognitive decline. To test such proposal, we used here machine learning approaches, previously successfully applied by our group in large epidemiological studies (Bravo-Merodio et al., 2019; COVIDSurg Collaborative et al., 2021) to model interactions between CR estimates (by proxy of education) and neurobiological features (measures of structural and functional frontoparietal connectivity and genetic variants in noradrenergic signalling) on cognitive outcomes in the ageing UK Biobank cohort.

METHODS

Participants. We used data from the UK Biobank, a prospective epidemiological cohort study with over 500,000 participants 40–69 years of age at recruitment (total of 502,505 participants), who underwent wide-ranging phenotypic and genotypic characterisation (Sudlow et al., 2015). For the purpose of the current study, we employed (1) genomic, (2) demographic data from the initial baseline visit at recruitment between 2006 and 2010, and (3) cognitive and (4) multimodal neuroimaging data acquired between 2014 and 2020 (the UK Biobank Brain Imaging Cohort; Miller et al., 2016; Alfaro-Almagro et al., 2018). At the time of

the study, multimodal neuroimaging data release included 48,561 participants, and so modelling was performed with these datasets only. All the analyses were conducted under the UK Biobank application numbers 29447 and 31224. All UK Biobank participants provided written informal consent in accordance with approved ethics protocols (REC reference number 11/NW/0382).

Demographic information and cognitive data. Basic demographic information including month and year of birth, sex and education was recorded as part of information acquired at recruitment or as part of touchscreen questionnaire completed during the baseline assessment visit. The full details of the touchscreen questionnaire and all procedures are provided on the UK Biobank website (https://biobank.ndph.ox.ac.uk/ukb). The month and year of birth alongside the date of attending imaging visit (MRI scanning session) were used to calculate participants age. For the purpose of the current study, we used sex as recorded at recruitment (information derived from central registry i.e., as recorded by NHS or if not available, self-reported information was used). Education variable was determined based on self-reported age of completing full time education. Most UK Biobank participants provided this information during baseline visit by answering a question "At what age did you complete your continuous full time education?", with those reporting they "Never went to school" having a 0 assigned. However, if this information was not captured during the baseline assessment (instances of a missing answer as opposed to participant answering "Prefer not to answer" (2760) or "Do not know" (3454)), it was gathered during the subsequent visits. Despite multiple attempts to acquire this information, we identified a large portion of participants with this information still missing, 46% (or 22,334 out of 48,705 participants) (Supplementary Table 1). After ascertaining that this information was Missing At Random (MAR) (Supplementary Table 1, Supplementary Figure 1) and given its pivotal role as proxy of

cognitive reserve, participants with missing education information were deleted from our analysis.

Cognitive assessment was completed on the same day as attending imaging visit and entailed completing several simple cognitive tests administered via touchscreen. The UK Biobank cognitive test battery consists of 13 different tests as detailed in the UK Biobank showcase (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100026). However, some of these tests were only used during piloting stage of the UK Biobank study and some tests added or modified for the subsequent recruitment phases and/or subsequent visits. Additional information about UK Biobank cognitive tests, alongside their validity and reliability assessment can be found in the previously published papers (Sudlow et al., 2015; Fawns-Ritchie and Deary, 2020). For the purpose of the current study we used a subset of measures indicative of cognitive performance on 10 different tests measuring memory, reasoning, executive function and processing speed: (1) reaction time (mean time in to correctly identify matches) derived from SNAP game, (2) numeric memory (maximum number of correctly recalled digits) based on short-term numeric memory test, (3) number of correct answers on the fluid intelligence test (fluid intelligence score) measure of cognitive outcome in the UK Biobank, (4) duration to complete numeric path in the version A of Trial Making test; (5) duration to complete alphanumeric path in the version B of Trial Making test, (6) number of correctly solved puzzles in the matrix patterns test, (7) number of correctly solved puzzles in the tower rearranging test, (8) symbol digit substitution - number of correct symbol-digit matches identified in the symbol digit substitution test, (9) number of word pairs correctly matched in the paired associate learning test, and (10) number of errors (incorrect card matches) in the pair matching test.

MRI data. The neuroimaging datasets available for the UK Biobank Brain Imaging Cohort (i.e., participants who completed imaging visit) include a set of imaging derived phenotypes (IDPs) i.e., various measures derived from different magnetic resonance imaging (MRI) modality-specific analyses (Alfaro-Almagro et al., 2018). The full information about the MRI data acquisition and data processing pipelines applied to the data provided by the UK Biobank is available online (https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/brain_mri.pdf) and in previously published primary UK Biobank methods papers (Miller et al., 2016; Alfaro-Almagro et al., 2018). For the purpose of the current study we selected a set of IDPs representing structural and functional fronto-parietal connectivity measures derived from diffusion MRI (dMRI) and resting state functional MRI (rs-fMRI) respectively. Specifically, we used dMRI IDPs reflecting microstructural properties of white matter pathways, which were derived from neurite orientation and dispersion imaging (NODDI) model and representing intra-cellular volume fraction (ICVF) as a measure of neurite packing density, and the orientation dispersion index (ODI) as a measure of dispersion of neurites (an estimate of fiber coherence) for three fronto-parietal white matter pathways, superior longitudinal fasciculus (SLF), inferior-fronto-occipital fasciculus (IFOF) and the forceps minor of the corpus callosum. In addition, we used rs-fMRI IDPs reflecting brain functional connectivity, which were generated by (1) carrying out group independent component analysis (ICA) parcellations with dimensionality (the number of distinct ICA components identified based on temporal patterns of spontaneous fluctuations in brain activity) set at 25, (2) removal of noise components resulting in 21 components representing separate resting state networks (functional nodes), and (3) estimating 21x21 partial correlation matrices, which represent connections (edges) between pairs of ICA components (nodes). For the purpose of the current study, we have selected 6 networks of interests (dorsal and ventral default mode network, left and right fronto-parietal network,

default mode network related network and executive control network; identified based on previously published papers (Smith et al., 2009; Laird et al., 2009; Shirer et al., 2012)) and connection (edges) between them were entered into the analysis (total of 15 different edges). The 25 components group-ICA spatial maps and connection edges are available online (http://www.fmrib.ox.ac.uk/ukbiobank).

Missing value imputation. Given education as proxy of cognitive reseve is one of the main variable studied, we first deleted all participants with missing information (22,334 participants) (Supplementary Table 1 and Supplementary Figure 1). From this dataset (26,227 participants and 39 variables spanning demographic data, cognitive tests and multimodal neuroimaging data), two different approaches were taken in order to deal with missing data information. First, and as mainly reported below, we performed a complete case analyses, deleting all patients with missing value information, after having seen no visible pattern in previous analysis and determining missing values were missing at random (MAR) (Supplementary Figure 2). The final dataset consisted of 12,076 particpants. The second approach consisted in imputing missing value data (20.2%) (Supplementray Figure 2A), using K-nearest neighbour imputation with three neighbours. Subsequently, the same analyses as with complete cases, were performed using imputed data, and are reported in Supplementary Materials. Further information, all codes and fugures are available from https://github.com/InFlamUOB/CognitiveReserve.

General cognitive ability score (GCAS) For a more actionable way in which to incorporate the cognitive test information in our model and following previous strategies, a measure of general cognitive ability was then calculated from the cognitive test scores (Fawns-Ritchie and Deary 2020). These was done using principal component analysis (PCA), saving scores based on the first principal component, which accounted for 30.9% of the variance following

previous reports (Fawns-Ritchie and Deary, 2020; Lyall et al., 2016). Eigenvalues and scree plots can be found in Supplementary Figure 3 and Supplementary Table 2. The calculated GCAS represents composite cognitive ability and as such was subsequently used as a measure of cognition.

GWAS derived SNPs. The genetic analyses were carried out based on genotype data derived from blood samples collected during baseline visit, in total UK Biobank includes genetic data for 488,377 participants. In accordance with UK Biobank protocols, following DNA extraction, the genotyping was performed using two different arrays, the UK BiLEVE Axiom array and the UK Biobank Axiom array. The full details of the genotyping, quality control and imputation have been previously published (Bycroft et al., 2018) and are available from the UK Biobank showcase (https://biobank.ndph.ox.ac.uk/ukb). In order to extract genetic features, we selected specific genetic variants which have been implicated in noradrenergic and dopaminergic neurotransmission. The final gene set, total of 154 genes (464,939 SNPs) was selected based on search performed using the Molecular Signature Database (MSigDB; https://www.gsea-msigdb.org/gsea/msigdb/), and published literature (Bralten et al., 2013; van Donkelaar et al., 2020) and can be found in in https://github.com/InFlamUOB/CognitiveReserve. In the analysis, we have included genes associated with both noradrenergic and dopaminergic genetic pathways as the two neurotransmitters share both signalling and metabolic pathways due to overlapping biosynthesis and the known promiscuity of some of the transporters and receptors (Carboni et al., 1990; Cornil and Ball, 2008; Wedemeyer et al., 2007). Thus, separating the two genetic pathways is highly problematic, if not impossible. These SNPs were then tested for associations with the previously mentioned metrics: cognitively enriched environments (education), cognitive outcomes (measures of performance on 10 cognitive tests) and fronto-

parietal connectivity (structural and functional connectivity IDPs), using genome-wide association analysis (GWAS). More specifically the tools PLINK (Purcell et al., 2007) for pre-processing and quality control and REGENIE (Mbatchou et al., 2021) for GWAS were used. REGENIE is a computationally efficient machine learning method that reduces the genotype space into local blocks (of 1000 variants) before performing association testing. After GWAS, chosen SNPs were filtered (LOG10P > 4.5), yielding 53 SNPs. Summary statistics of selected SNPs can be seen in https://github.com/InFlamUOB/CognitiveReserve. In addition, to control for potential confounding effects in our analysis, we included SNPs e3e3, e3e4 and e4e4 from apolipoprotein-E (APOE), which is known to be a risk factor for dementia (Heffernan et al., 2016).

The distribution of all final variables introduced in the model encompassing (1) GWAS derived SNPs (2) demographics (age and sex) (3) general cognitive ability score (GCAS) from cognitive tests (4) imaging derived penotypes representing structure and functional front-parietal connectivity (FPC IDPs) and (5) cognitive reserve information using education as proxy can be seen in Supplementary Figure 5.

Statistical analysis. All analyses were performed in R version 4.2.0

Modelling

As our objective is to model interactions between neurobiological features (both genetic and connectivity measures) and cognitive reserve (by proxy of education) on cognitive outcomes, we built regression models using general cognitive ability score (GCAS) as outcome. Using *workflowsets* and *tidymodels*, 5 different algorithms (generalized linear model (g*lm* package), LASSO (Tibshirani, 1996) (*glmnet* package*)*, a single-hidden-layer neural network (*nnet* package*)* random forest (*ranger* package) and xgboost (*xgboost* package), were fitted to 14

different combinations of the features above extracted. More precisely, 5 different algorithms were run on data comprising: SNPs, FPC IDPs, AgeSex, SNPs x FPC IDPs, SNPs x CR (education as proxy), SNPs x AgeSex, FPC IDPs x CR (education as proxy), AgeSex x CR (education as proxy), SNPs x FPC IDPs x CR (education as proxy), SNPs x FPC IDPs x AgeSex, SNPs x AgeSex x CR (education as proxy), FPC IDPs x AgeSex x CR (education as proxy), SNPs x FPC IDPs x AgeSex x CR (education as proxy) information. In each algorithm and feature combination dataset (69 in total) (Figure 1), hyperparameters were tuned through *grid.search* and 100 cross-validation resamples fitted per model. To cut down on time, *tune_race_anova* (Kuhn, 2014) was used which eliminates tuning parameter combinations that are unlikely to be the best results. To do so*, tune_race_anova* uses a repeated measure ANOVA model just after an initial number of resamples have been evaluated. Performance was evaluated through Mean Absolute Error (MAE), where the sum of absolute errors between predicted GCAS and true GCAS is divided by the sample size.

Variable importance

To better understand the behaviour of the models, variable importance of each of the 69 different algorithm and variable combinations was assessed using 4 different feature importance methods (Elhabr, 2020). These include model-agnostic algorithms such as the explainable AI package *DALEX* (Biecek, 2018), shapley values (Sundararajan and Najmi, 2020) and permutation related variable importance or model-specific algorithms such as model related variable importance. The loss function in variable importance computations for the model-agnostic methods was minimization of the root mean-sqaured error (RMSE). The top 9 ranking features in each of the 69 different models generated was then visualized in a heatmap and those scoring high importance (>3) are then selected and followed through for mediation analysis.

Mediation

Finally, to assess a possible relationship between cognitive reserve (by proxy of education) on cognitive outcomes, we tested those features that came up as highly important in our models through mediation analysis (VanderWeele, 2016). Mediation analysis was used to assess the magnitude of the data-driven pathways and mechanisms selected as relevant in a data-driven way and how they may affect general cognitive ability score and CR relationship. Mediation was performed using libraries *psych* and *mediation* where all 38 consistently important variables for GCAS prediction (top 3 important variables in each of the 69 models) were studied. Mediation diagram for all 38 variables can be seen in in https://github.com/InFlamUOB/CognitiveReserve with the mediation effect bootstrapped 1000 times.

RESULTS

Pre-processing

From an initial cohort of 502,505 participants, 48,561 with imaging data are followed through. From these, 26,227 had complete education information. Missing values assessment can be seen in (Supplementary Figure 1 and 2), with remaining missing values (20.2%) both imputed using k-nearest neighbours (Supplementary Figure 3) and deleted (complete-case analysis) as reported below. Taking advantage of genotypic data, SNPs related to noradrenergic and dopaminergic neurotransmission were extracted as well as APOE alleles e3e3, e3e4 and e4e4 given their close relationship to cognitive decline and dementia (Heffernan et al., 2016; Veldsman et al., 2020). In total 12,076 participants with 25 features associated to brain imaging data, 56 genetic data, 2 demographics and education were selected with their distribution available in Supplementary Figure 5.

Modelling

To understand what data better captured general cognitive ability, different models were built predicting GCAS using both different feature combinations and modelling strategies (Figure 1). More specifically, generating all different feature combinations between brain imaging, genetic, demographics and education yielded 14 different datasets as seen in Methods. All these datasets were then trained using 4 distinct machine learning algorithms combining both non-linear and linear methods (Random Forest, Neural Nets, LASSO and GLM) creating 69 different models. After splitting our dataset into training (5/6) and testing (1/6) data, models were trained, with hyperparameter tuning and resampling (100 fold-cross validation).

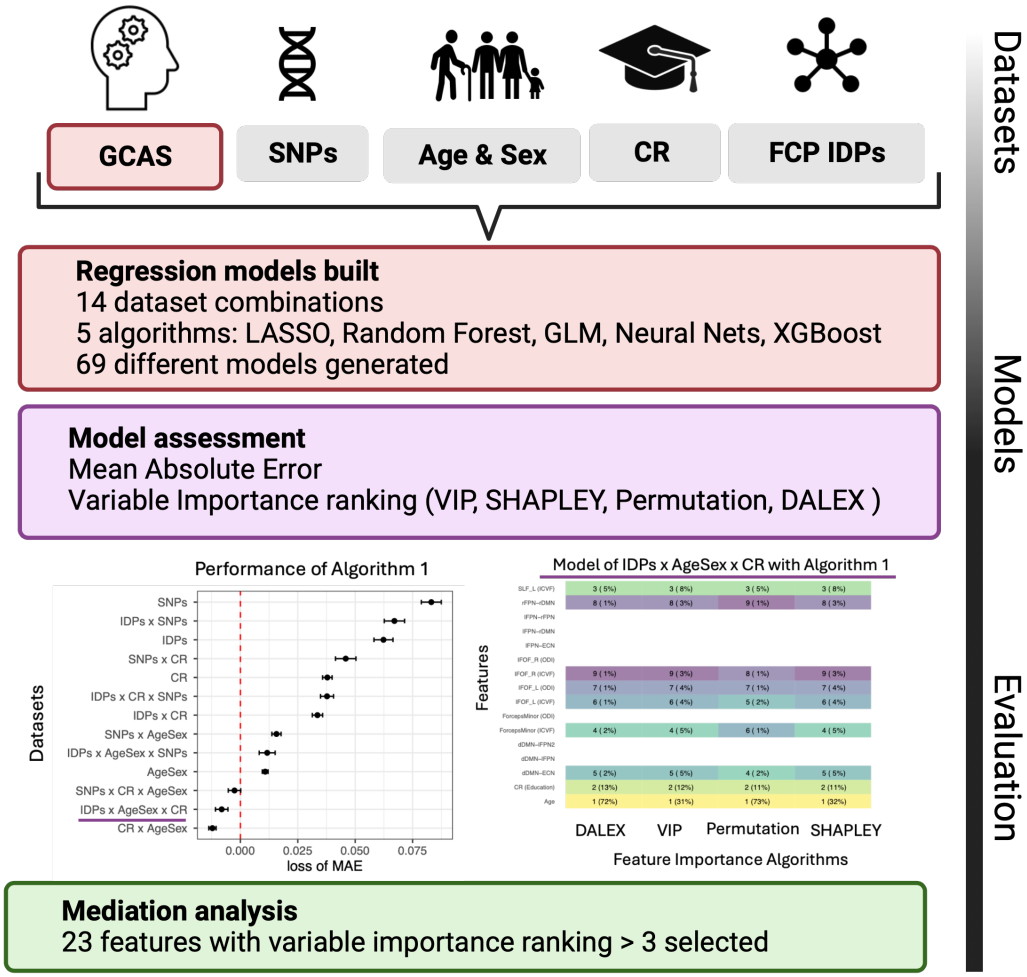


Figure 1. Study Design. We explored the interplay between genetics and environmental factors contributing to cognitive reserve by testing regression models prediciting general cognitive ability score (GCAS) as outcome. By splitting features into 4 datasets: GWAS derived SNPs, AgeXSex, CR with Education as proxy and imaging derived penotypes representing structure and functional front-parietal connectivity (FPC IDPs), we generated 14 different dataset combinations. These 14 different feature combimations were then modelled with 5 different algorithms (Least Absolute Shrinkage and Selection Operator (LASSO), Generalized Linear Models, XGBoost, Random Forest and Neural Nets), generating 64 different models. Behaviour of these models was then evaluated through performance, using mean absolute error and feature importance. Given the disparity in modelling algorithms, different feature importance algorithms were also used (DALEX, model dependent variable importance assessment, permutation and SHAPLEY), and results were summarised as ranking scores. All features with a ranking of 3 or less per dataset and algorithm were then studied as mediators between CR and GCAS.

Final performance was assessed in the test set (Figure 2) and feature importance through 4 metrics (vip, permutation, shapley and DALEX) for all models is reported (Figure 2).

A combination of imaging, demographic and education data was seen to be the best performing model in all algorithms, with these feature combinations yielding better performance than combining all features together (reference line in Figure 2). To better understand feature importance per model and algorithm, heatmaps were used to visualize results, with information on the models generated form imaging, demographic and education data seen in Figure 3 as an example. In this heatmap, the ranking of the top 9 features per feature importance methods applied to each model is reported, with Age and Education scoring first and second in all algorithms and methods but less consensus seen in the ranking of imaging data, with varying agreement between algorithms and feature importance methods. Results from the other models can be found in: https://github.com/InFlamUOB/CognitiveReserve

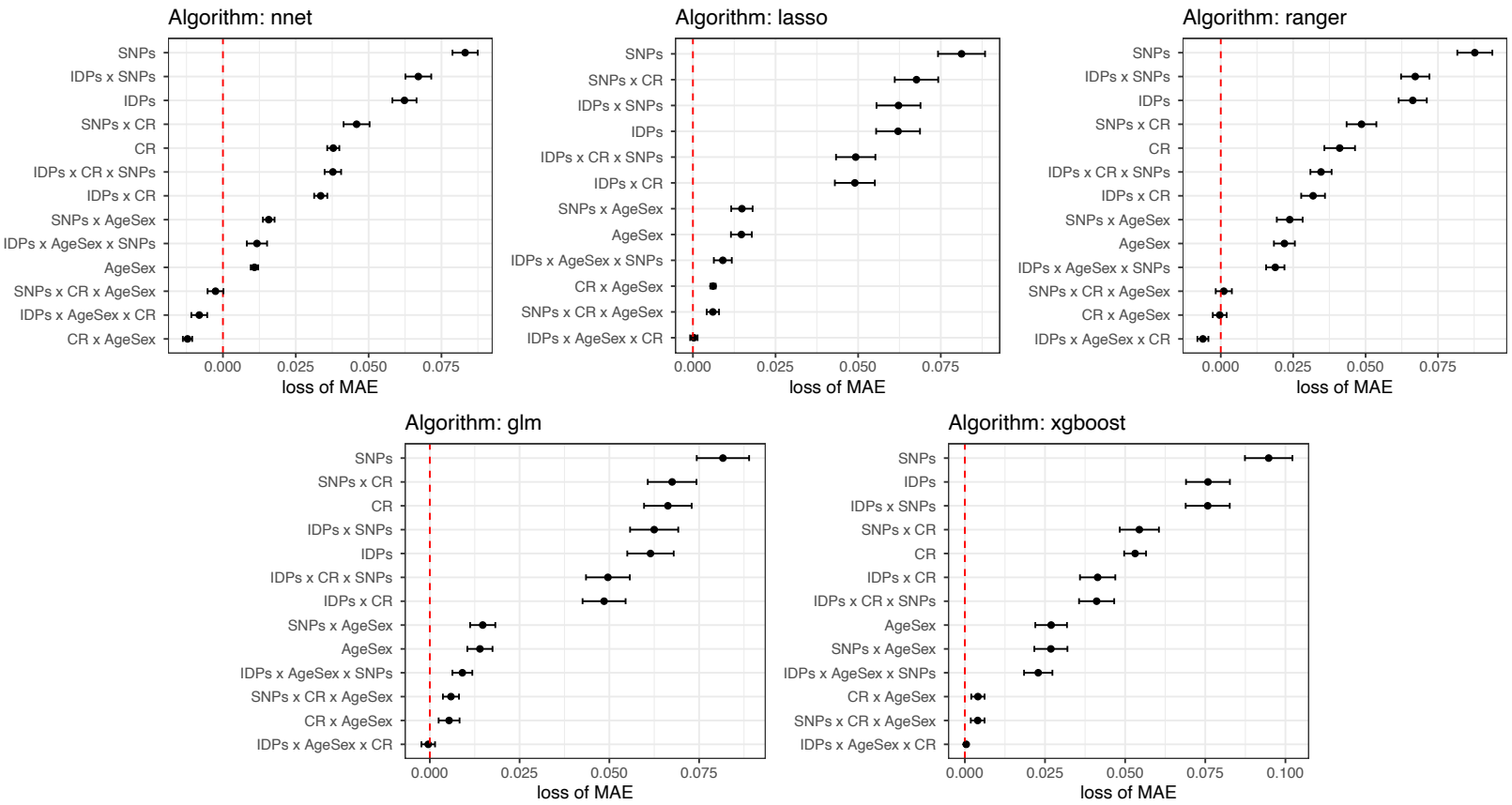


Figure 2. Mean Absolute Error (MAE) of model combinations of AgeXSex, front-parietal connectivity imaging derived penotypes (FPC IDPs) and SNPs data for each of the different algorithms (Neural Nets, LASSO, Random Forest, XGBoost and GLM), using the model with all datasets included as reference value (red line). Error bars correspond to the 100 fold cross-validated performance results. In total, 64 models were analyzed (with unique feature and algorithm combination). As a general trend, SNPs possess the least prediction ability, with the highest mean average error across all algorithms (higher loss of MAE with respect to baseline 0 (which is all datasets together). In contrast, including AgeXSex x CR guarantees a better performance across all datasets, with IDPs x AgeSex x CR generally yielding the best performance. All hyperparamters were tuned using tune race_anova with best perfroming hyperparameter combinations used and test set performance is reported here.

Mediation

Finally, wanting to better understand the relationship between education and cognitive outcome, we proceeded to explore the top important features extracted from each model (ranking >3) (23 different features selected) in a mediation model. More specifically, the direct effects of CR (with education as proxy) on general cognitive ability score partialling out the effect of each of these features univariately was evaluated with the main effects reported in Table 1. Following that seen as best model (FPC IDPs, CR and Age and Sex features), those variables seen to have a significant mediation effect were mainly Age and rFPN-rDMN as well as dDMN-ECN (Figure 4). The resuls based on the imputed data are reported in the Supplementary Figure 4. Age and rFPN-rDMN were found as significant medaiors in both analyses i.e., using only complete cases and imputed data, thus ascertaining the robustness of finidngs.

Code Availability

All code and figures can be found in: https://github.com/InFlamUOB/CognitiveReserve

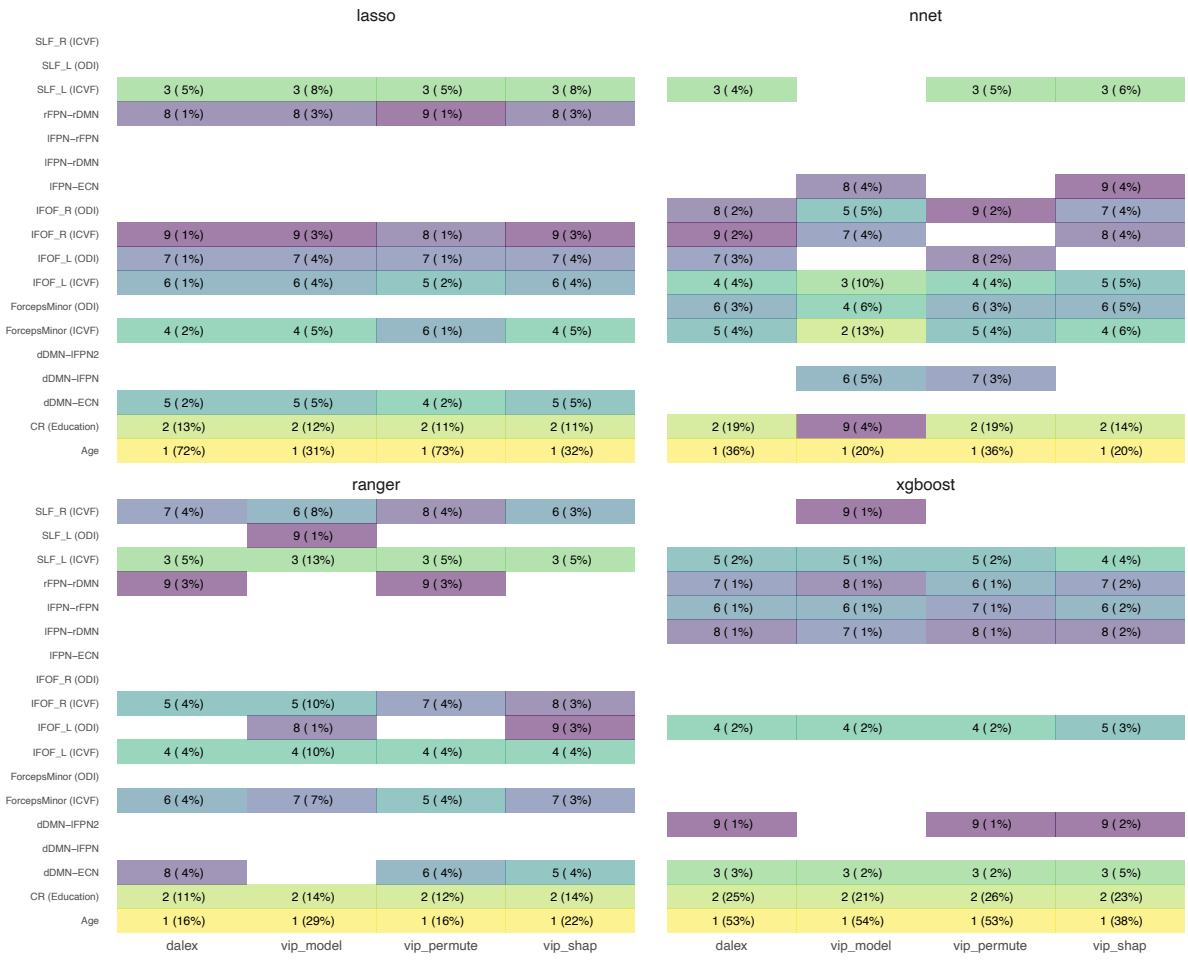


Figure 3. Heatmap of feature importance for the 4 out of the 5 different algorithms (LASSO and GLM as both linear models, are here only represented as LASSO results) and data comprising features from AgeXSex, CR with education as proxy and front-parietal connectivity imaging derived penotypes (FPC IDPs). Most important features are ranked using 4 different feature importance algorithms (*DALEX*, model specific variable importance (*VIP*) , permutation and Shapley values), with most important being 1 (yellow) and least 9 (purple). Only the 9 most important features for each models are here reported. All model-agnostic methods (permutation, DALEX and Shapley) had variable importance assessed though root mean-squared error (RMSE) and the absolute values of importance were assessed as ranking. In brackets, the normalized absolute importance value is reported as percentage.

Table 1. Mediation results using *psych* package. Selected features (most important variables (>3 ranking)) were included in a model as mediators with CR as predictor and General Cognitive Ability Score (GCAS) as outcome. Bootstrapped confidence intervals for the indirect effect reported as lower CI (2.5%) and upper (97.5%). In red significant mediators (95% CI not crossing 0).

Mediator	Effect of CR on mediator (a)	Effect of mediator (b)	Indirect effect (ab)	95% CI	
				Lower	Upper
rs17046359(SYT1)	0.0016 (0.0013)	-0.0271 (0.061)	-1e-04 (1e-04)	-4e-04	2e-04
Age	-0.0865 (0.0091)	-0.3773 (0.0083)	0.0326 (0.0036)	0.0256	0.0398
rs1872691(BRD7-ADCY7)	0.0013 (0.0042)	-0.0465 (0.0196)	-1e-04 (2e-04)	-5e-04	4e-04
rs62465540	-3e-04 (4e-04)	0.1883 (0.2263)	-1e-04 (1e-04)	-3e-04	1e-04
rs10224002(PRKAG2)	-0.0016 (0.0026)	0.002 (0.032)	0 (1e-04)	-2e-04	2e-04
SLF_L (ICVF)	0.0062 (0.0091)	0.165 (0.0088)	0.001 (0.0015)	-0.0019	0.0041
IFOF_L (ICVF)	0.0161 (0.0091)	0.1562 (0.0089)	0.0026 (0.0014)	-2e-04	0.0054
rFPN-rDMN	-0.0362 (0.0091)	-0.0502 (0.009)	0.0018 (6e-04)	8e-04	0.003
IFOF_L (ODI)	-0.0057 (0.0091)	-0.022 (0.009)	1e-04 (2e-04)	-3e-04	6e-04
IFOF_R (ICVF)	0.0143 (0.0091)	0.1506 (0.0089)	0.0022 (0.0014)	-5e-04	0.0049
rs7786782(PRKAG2)	-0.0017 (0.0045)	-0.0201 (0.018)	0 (1e-04)	-2e-04	3e-04
rs10254101(PRKAG2)	-0.0021 (0.0045)	-0.0088 (0.0182)	0 (1e-04)	-2e-04	3e-04
dDMN-ECN	-0.018 (0.0091)	0.0639 (0.009)	-0.0011 (6e-04)	-0.0024	0
ForcepsMinor (ICVF)	0.0106 (0.0091)	0.1514 (0.0089)	0.0016 (0.0013)	-0.001	0.0042
lFPN-ECN	-0.0083 (0.0091)	-0.0312 (0.009)	2e-04 (3e-04)	-3e-04	9e-04
rs2888788(PRKAG2)	-0.002 (0.0045)	-0.0313 (0.018)	1e-04 (2e-04)	-3e-04	4e-04
Sex	0.0078 (0.0045)	-0.044 (0.018)	-3e-04 (3e-04)	-9e-04	1e-04
rs3813755(ADCY7)	0.0012 (0.0042)	-0.0433 (0.0196)	-1e-04 (2e-04)	-5e-04	3e-04
rs6947064(PRKAG2)	-0.0033 (0.0045)	-0.0194 (0.018)	1e-04 (1e-04)	-2e-04	4e-04
rs118186402(GNA14)	6e-04 (0.0014)	-0.1709 (0.059)	-1e-04 (3e-04)	-7e-04	4e-04
e3e3(APOE)	-0.0061 (0.0045)	0.0236 (0.0182)	-1e-04 (2e-04)	-6e-04	1e-04
rs796354214	-0.0015 (0.0038)	0.013 (0.0216)	0 (1e-04)	-2e-04	2e-04
rs62468660	-2e-04 (3e-04)	0.1498 (0.2392)	0.1695 (0.009)	0 (1e-04)	-2e-04

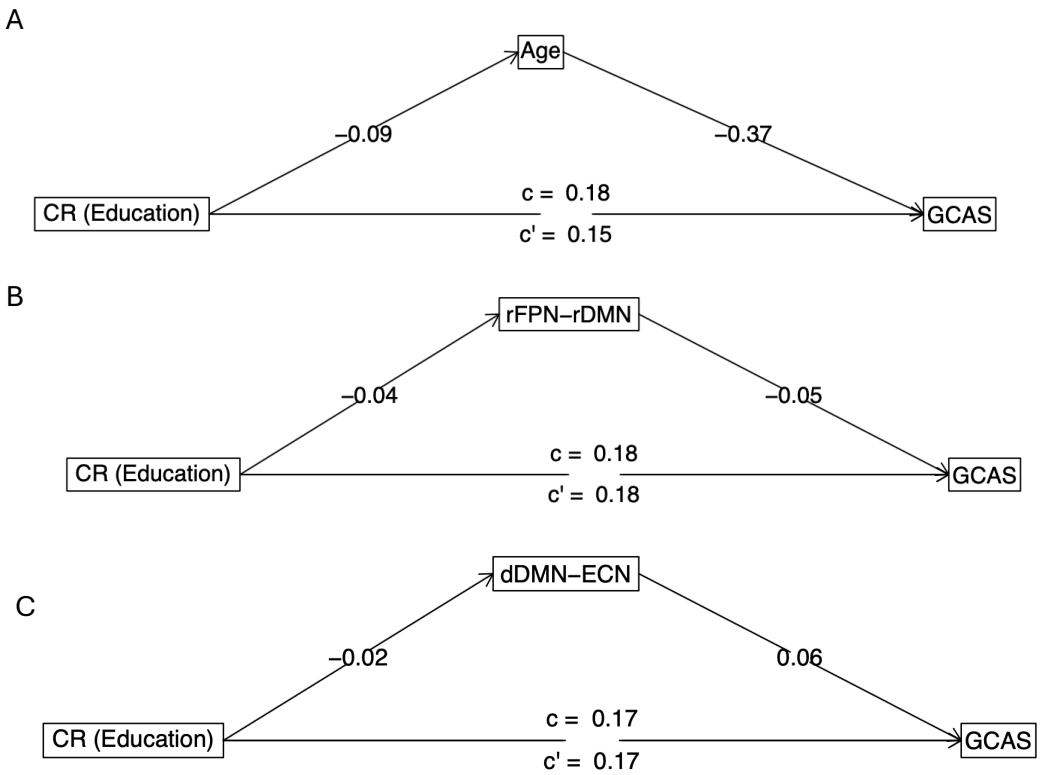


Figure 4. Mediation diagram as extracted from mediate.diagram function from the psych package. Significant mediators (95% CI not crossing 0) Age, rFPN-rDMN and dDMN-ECN are reported .

DISCUSSION

Using robust machine learning methods and mediation analysis applied to data from UK Biobank Brain Imaging study (Miller et al., 2016), we tested here combinatorial effects of neurobiological features (genetic variants in noradrenergic signalling, measures of structural and functional frontoparietal connectivity) and environmental influences (cognitively enriched environments by proxy of education) on cognitive outcomes in the UK Biobank ageing cohort. Following building several different models with initial large number of features, and subsequent variable importance testing, we selected a small number of features, related to genetic variance, fronto-parietal connectivity alongside sex and age, for subsequent analysis using mediation testing. A variety of machine learning algorithms and

feature importance methods were used given that there is no algorithm able to fit all data perfectly ("No Free Lunch Theorem"(Belkin et al., 2019)). Therefore, testing more than one algorithm and more than one approach to interpret feature importance provides us with the possibility of acknowledging many more possible underlying relationships. Also feature importance was especially relevant for black box algorithms such as neural nets and random forest where behaviour of the model is more difficult to grasp. Based on these analyses we show that (1) interactions between education and neurobiological variables more fully explain general cognitive performance on multiple tests (general cognitive ability score) than either factor alone, and (2) among the examined neurobiological features measures of functional fronto-parietal connectivity are the strongest mediators of the effect of education on cognitive outcomes.

Our main findings point to two fronto-parietal resting state functional connectivity measures (rFPN-rDMN and dDMN-ECN, Figure 4) representing, (i) connections between the default mode network and the right fronto-parietal network, and (ii) connections between the default mode network and the executive control network as significant mediators between cognitively enriched environments and cognitive outcomes in UK Biobank ageing population sample. This suggests that functional connectivity between several neural networks, including right fronto-parietal network, might be sensitive to the effects of education (proxy of exposure to enriched environments across lifespan) on cognition in middle and old age. Thus, alongside our previous studies (Shalev et al., 2020; Brosnan et al., 2022), we provide here evidence in support of Robertson's theory of CR (Robertson 2013, 2014). Robertson's theory stipulates that at the neural level the lifelong exposure to enriched environments strengthens the right hemisphere fronto-parietal networks, which in turn offsets the symptoms of cognitive ageing. Our previous work (Brosnan

et al., 2022; Shalev et al., 2020) and that of others (van Loenhoud et al. 2017), indeed point to the right hemispheric fronto-parietal networks as neural substrates of CR. By contrast here we linked the phenomenon of CR to not only the right fronto-perietal network but also the default mode network and the executive control network. The Robertson's theory proposes that exposure to enriched environments acts via core cognitive processes, including alertness, sustained attention, and awareness which indeed are known to be subserved by the right fronto-parietal networks (e.g., Beck et al., 2001; Naghavi and Nyberg, 2005; Hester et al., 2005; Singh-Curry and Husain, 2009; Stuss, 2011). However, both the executive control and default mode networks implicated by our current study are also functionally linked to the cognitive processes described by Robertson's. Furthermore, both default mode network and executive control network are known to be modulated by the locus coeruleus-noradrenergic (LC-NE) system (Bär et al., 2016; Minzenberg et al., 2018; Robbins & Arnsten, 2009; Suttkus et al., 2021), which is pivotal to Robertson's theory.

It should be also noted that while the mentioned above studies, supporting Robertson's theory, only examined the beneficial effects of enriched environments on structural measures within fronto-parietal networks (van Loenhoud et al. 2017; Shalev et al., 2020; Brosnan et al., 2022;), the current study links the phenomenon of CR to measures of functional connectivity. Several previous studies in neurotypical older adults linked measures of functional connectivity to offsetting effects of cognitive reserve either on global cognition or memory performance (e.g., Varela-Lopez et al., 2022; Boyle et al., 2023; Franzmeier et al., 2018; Fleck et al., 2019). However, most of these studies either explored connectivity measures derived from a task-based fMRI and/or the analysis were primarily restricted to measures of within-network connectivity, and additionally the previous findings were limited by relatively small sample sizes (less than 100 participants). By contrast we explored here

between networks connectivity in a large cross-sectional UK Biobank cohort. But what's most striking about our findings is that they point to the default mode network as the key player in between networks interactions underlying cognitive reserve. In the context of ageing, the default mode network is perhaps the most studied of the resting state networks (Jiang et al., 2020; Staffaroni et al., 2018) with functional changes linked to cognitive decline and dementia (Vidal-Peneiro et al., 2014; Andrews-Hanna et al., 2007; Damoiseaux et al., 2008; Bluhm et al., 2008; Mevel et al., 2011; Hafkemeijer et al., 2012; Grieder et al., 2018). In addition, when examining a small sample of Alzheimer's patients, Bozzali and colleagues (2015) previously found evidence that cognitive reserve modulate connectivity within the default mode network. Finally, our results expand on newly published findings linking changes in connectivity between the default mode network and frontoparietal networks to cognitive decline in elderly (Koshino et al., 2023) and to the effects of beta-amyloid on cognitive status in Alzheimer's patients (Zhukovsky et al., 2023).

To our knowledge, only one previous study examined underpinnings of the effects of CR on cognitive outcomes in the UK Biobank cohort (Jin et al., 2023). However, the scope, research objectives and methods employed by Jin et al (2023) and here are very different. Jin and colleagues specifically conducted a series of studies systematically exploring the effects of different CR proxies (education, leisure activities, fluid intelligence, social interactions, and physical activity) on the relationship between structural brain measures (global volume, regional volumetric measures and cortical thickness within brain areas know to deteriorate in dementia patients) and cognitive outcomes indexed by performance on tests within 4 separate cognitive domains. The main aim of their research was to explore the reported in literature inconsistencies concerning the moderating effects of various CR proxies on the links between brain structure and cognitive abilities in ageing population i.e., aim motivated by the

premise that CR explains the discrepancy between accumulated brain damage and observed cognitive performance in accordance with classical theory of cognitive reserve as proposed by Stern and based on his early work in Alzheimer's patients (Stern et al., 1992; Stern, 2012). By contrast our analyses explored the hypothesis that CR represented by proxy of education strengthen the right fronto-parietal networks to offset age-related cognitive decline. Thus, the only linked between Jin et al (2023) and our study, conducted using data from UK Biobank cohort, is that both provide support for the notion of beneficial effects of CR on cognitive ageing.

One of the objectives of the current study was to explore a premise that an interplay between accumulative effects of cognitively enriched environments and genetic variance in noradrenergic signalling on frontoparietal neural networks might offset age-related cognitive decline. While previous studies indeed linked genetic variations in neurotransmitter signalling to heterogeneity in cognitive ageing (Lindenberger et al., 2008; Nyberg et al., 2014; Papenberg et al., 2015a,b), we have not found any evidence of the mediating effects of genetic variability in noradrenergic signalling on interplay between cognitively enriched environments and cognitive outcomes. It should be noted that the previous evidence comes from hypothesis driven genetic association studies using single-nucleotide polymorphisms (SNPs). By contrast we employed here GWAS based approach to find genetic variants associated with cognitively enriched environments (education), cognitive outcomes and fronto-parietal connectivity (structural and functional). While the GWAS analyses were notably restricted to the preselected gene set, 154 genes totalling 464,939 SNPs, the negative findings might not be that surprising considering a relatively small sample comparing to other GWAS studies examining genetic underpinnings of education and cognition (e.g., Okbay et al., 2016; Davies et al., 2016). Thus, the size of study sample constitutes a limitation when

exploring genetic influences. The negative findings might also reflect overall shortcomings (anoter limitation) of the taken here approach to testing Robertson's theory, specifically that genetic variance in noradrenergic signalling does not capture and/or does not reflect cumulative influences of cognitively enriched environments as per Robertson's theory of cognitive reserve. It should be also noted that the beneficial effects of noradrenergic signalling in the context of both cognitive decline and progression of Alzheimer's disease have been recently questioned by many (for comprehensive review see Mather, 2021). Finally, the scope and size of UK Biobank data resources are unprecedented, however similarly to other large scale epidemiological datasets, UK Biobank includes large proportion of incomplete cases, cases with missing data. Here, out of initial sample of 48,561 participants, only 12,076 complete cases were found (i.e., including all brain imaging data, cognitive, genetic, demographic and education variables). The missing cases cause potential limitations due to loss of statistical power (e.g. in GWAS as discussed above) and by contributing to risk of bias affecting findings. To address that we repeated all analyses using imputed data. Importanlty, we again found the measures of fronto-parietal connectivity to be the strongest mediators of the effect of CR on cognitive outcomes.

In summary, the presented here data not only add to our understanding of the neurobiological underpinnings of CR but also the role of between network connectivity and the default mode network in cognitive ageing. By employing comprehensive machine learning methods, we have demonstrated that between networks functional fronto-parietal connectivity is a strong mediator of the effect of cognitive reserve on cognitive outcomes in ageing population. Our findings provide additional support for Robertson's theory of CR (Robertson, 2013, 2014), and recent finding in Alzheimer's patients linking between networks frontoparietal connectivity to offsetting effects of CR against amyloid burden (Zhukovsky et

al., 2023). While on one hand the limitation of the current study is that we only examined between network connections, on the other hand several theories of cognitive ageing explicitly address changes in functional correlations between functional networks associated with either preservation or decline in cognitive performance in ageing. Specifically, it has been suggested that processes such as compensation and dedifferentiation both might be accompanied by stronger correlations among functionally unrelated networks or weaker anti-correlations among competing networks (Malagurski et al., 2020a,b; Deery et al., 2023). In this context the default mode network is of particular interest as the existing evidence points to anti-correlations between the default mode network and various other networks including fronto-parietal networks at rest and during task performance in young healthy participants (Greicius et al., 2003; Fox et al., 2005). Strikingly, it has been also demonstrated that in elderly participants the default mode network is significantly less deactivated during task performance and that this diminished suppression might underly poorer cognitive performance (Damoiseaux et al., 2008; Persson et al., 2007; Ferreira and Busatto, 2013). Thus, future research should address whether the offsetting effects of exposure to enriched environments on age-related cognitive decline translate into changes in co-activations and anti-correlations between the default mode networks and task-related fronto-parietal networks during various cognitive task performance.

ACKNOWLEDGMENTS: This research has been conducted using the UK Biobank Resource under Application Numbers 29447 and 31224.

FUNDING: This work was supported by The Royal Society International Exchanges Award (IES\R2\181100 to MC) and the Wellcome Trust Institutional Strategic Support Fund critical

data award (204846/Z/16/Z to MC). JAW acknowledges support from support from the MRC HDR UK (HDRUK/CFC/01), an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Medical Research Council or the Department of Health.

DISCLOSURES: Jackie A. Williams is currently an employee of Eisai, Inc. Eisai, Inc had no role in funding or design of this study.

REFERENCES:

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.R., Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D.C., Zhang, H., Dragonu, I., Matthews, P.M., Miller, K.L., Smith, S.M., 2018. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage 166, 400-424.

Andrews-Hanna, J.R., Snyder, A.Z., Vincent, J.L., Lustig, C., Head, D., Raichle, M.E., Buckner, R.L., 2007. Disruption of large-scale brain systems in advanced aging. Neuron 56(5), 924-935.

Avery, M.C., Krichmar, J.L., 2017. Neuromodulatory Systems and Their Interactions: A Review of Models, Theories, and Experiments. Front Neural Circuits 11, 108.

Bar, K.J., de la Cruz, F., Schumann, A., Koehler, S., Sauer, H., Critchley, H., Wagner, G., 2016. Functional connectivity and network analysis of midbrain and brainstem nuclei. Neuroimage 134, 53-63.

Beck, D.M., Rees, G., Frith, C.D., Lavie, N., 2001. Neural correlates of change detection and change blindness. Nat Neurosci 4(6), 645-650.

Belkin, M., Hsu, D., Ma, S., Mandal, S., 2019. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences 116(32), 15849-15854.

Biecek, P., 2018. DALEX: Explainers for Complex Predictive Models in R. J. Mach. Learn. Res. 19, 84:81-84:85.

Bluhm, R.L., Osuch, E.A., Lanius, R.A., Boksman, K., Neufeld, R.W., Theberge, J., Williamson, P., 2008. Default mode network connectivity: effects of age, sex, and analytic approach. Neuroreport 19(8), 887-891.

Boyle, R., Connaughton, M., McGlinchey, E., Knight, S.P., De Looze, C., Carey, D., Stern, Y., Robertson, I.H., Kenny, R.A., Whelan, R., 2023. Connectome-based predictive modelling of cognitive reserve using task-based functional connectivity. Eur J Neurosci 57(3), 490-510.

Bozzali, M., Dowling, C., Serra, L., Spano, B., Torso, M., Marra, C., Castelli, D., Dowell, N.G., Koch, G., Caltagirone, C., Cercignani, M., 2015. The impact of cognitive reserve on brain functional connectivity in Alzheimer's disease. J Alzheimers Dis 44(1), 243-250.

Bralten, J., Franke, B., Waldman, I., Rommelse, N., Hartman, C., Asherson, P., Banaschewski, T., Ebstein, R.P., Gill, M., Miranda, A., Oades, R.D., Roeyers, H., Rothenberger, A., Sergeant, J.A., Oosterlaan, J., Sonuga-Barke, E., Steinhausen, H.C., Faraone, S.V., Buitelaar, J.K., Arias-Vasquez, A., 2013. Candidate genetic pathways for attention-deficit/hyperactivity disorder (ADHD) show association to hyperactive/impulsive symptoms in children with ADHD. J Am Acad Child Adolesc Psychiatry 52(11), 1204-1212 e1201.

Bravo-Merodio, L., Acharjee, A., Hazeldine, J., Bentley, C., Foster, M., Gkoutos, G.V., Lord, J.M., 2019. Machine learning for the detection of early immunological markers as predictors of multi-organ dysfunction. Sci Data 6(1), 328.

Bravo-Merodio, L., Williams, J.A., Gkoutos, G.V., Acharjee, A., 2019. -Omics biomarker identification pipeline for translational medicine. J Transl Med 17(1), 155.

Brosnan, M.B., Arvaneh, M., Harty, S., Maguire, T., O'Connell, R., Robertson, I.H., Dockree, P.M., 2018. Prefrontal Modulation of Visual Processing and Sustained Attention in Aging, a tDCS-EEG Coregistration Approach. J Cogn Neurosci 30(11), 1630-1645.

Brosnan, M.B., Shalev, N., Ramduny, J., Sotiropoulos, S.N., Chechlacz, M., 2022. Right fronto-parietal networks mediate the neurocognitive benefits of enriched environments. Brain Commun 4(2), fcac080.

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.
Cabeza, R., Albert, M., Belleville, S., Craik, F.I.M., Duarte, A., Grady, C.L., Lindenberger, U., Nyberg, L., Park, D.C., Reuter-Lorenz, P.A., Rugg, M.D., Steffener, J., Rajah, M.N., 2018. Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. Nat Rev Neurosci 19(11), 701-710.

Carboni, E., Tanda, G.L., Frau, R., Di Chiara, G., 1990. Blockade of the noradrenaline carrier increases extracellular dopamine concentrations in the prefrontal cortex: evidence that dopamine is taken up in vivo by noradrenergic terminals. J Neurochem 55(3), 1067-1070.

Cornil, C.A., Ball, G.F., 2008. Interplay among catecholamine systems: dopamine binds to alpha2-adrenergic receptors in birds and mammals. J Comp Neurol 511(5), 610-627.
Fawns-Ritchie, C., Deary, I.J., 2020. Reliability and validity of the UK Biobank cognitive tests. PLoS One 15(4), e0231627.

COVID Collaborative, 2021. Machine learning risk prediction of mortality for patients undergoing surgery with perioperative SARS-CoV-2: the COVIDSurg mortality score. Br J Surg 108(11), 1274-1292.

Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., et al., 2015. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). Mol Psychiatry 20(2), 183-192.

Davies, G., Marioni, R.E., Liewald, D.C., Hill, W.D., Hagenaars, S.P., Harris, S.E., Ritchie, S.J., Luciano, M., Fawns-Ritchie, C., Lyall, D., Cullen, B., Cox, S.R., Hayward, C., Porteous, D.J., Evans, J., McIntosh, A.M., Gallacher, J., Craddock, N., Pell, J.P., Smith, D.J., Gale, C.R., Deary, I.J., 2016. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Mol Psychiatry 21(6), 758-767.

Damoiseaux, J.S., Beckmann, C.F., Arigita, E.J., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Rombouts, S.A., 2008. Reduced resting-state brain activity in the "default network" in normal aging. Cereb Cortex 18(8), 1856-1864.

Deery, H.A., Di Paolo, R., Moran, C., Egan, G.F., Jamadar, S.D., 2023. The older adult brain is less modular, more integrated, and less efficient at rest: A systematic review of large-scale resting-state functional brain networks in aging. Psychophysiology 60(1), e14159.

ElHabr, T., 2020 (r on Tony ElHabr 2020. 'Comparing Variable Importance Functions (For Modeling) | R-Bloggers'. 13 July 2020. https://www.r-bloggers.com/2020/07/comparing-variable-importance-functions-for-modeling/)

Ferreira, L.K., Busatto, G.F., 2013. Resting-state functional connectivity in normal brain aging. Neurosci Biobehav Rev 37(3), 384-400.

Fleck, J.I., Arnold, M., Dykstra, B., Casario, K., Douglas, E., Morris, O., 2019. Distinct Functional Connectivity Patterns Are Associated With Social and Cognitive Lifestyle Factors: Pathways to Cognitive Reserve. Front Aging Neurosci 11, 310.
Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102(27), 9673-9678.

Franzmeier, N., Hartmann, J., Taylor, A.N.W., Araque-Caballero, M.A., Simon-Vermot, L., Kambeitz-Ilankovic, L., Burger, K., Catak, C., Janowitz, D., Muller, C., Ertl-Wagner, B., Stahl, R., Dichgans, M., Duering, M., Ewers, M., 2018. The left frontal cortex supports reserve in aging by enhancing functional network efficiency. Alzheimers Res Ther 10(1), 28.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A 100(1), 253-258.

Grieder, M., Wang, D.J.J., Dierks, T., Wahlund, L.O., Jann, K., 2018. Default Mode Network Complexity and Cognitive Decline in Mild Alzheimer's Disease. Front Neurosci 12, 770.

Hafkemeijer, A., van der Grond, J., Rombouts, S.A., 2012. Imaging the default mode network in aging and dementia. Biochim Biophys Acta 1822(3), 431-441.

Harris, S.E., Deary, I.J., 2011. The genetics of cognitive ability and cognitive ageing in healthy older people. Trends Cogn Sci 15(9), 388-394.

Hayden, K.M., Reed, B.R., Manly, J.J., Tommet, D., Pietrzak, R.H., Chelune, G.J., Yang, F.M., Revell, A.J., Bennett, D.A., Jones, R.N., 2011. Cognitive decline in the elderly: an analysis of population heterogeneity. Age Ageing 40(6), 684-689.

Heffernan, A.L., Chidgey, C., Peng, P., Masters, C.L., Roberts, B.R., 2016. The Neurobiology and Age-Related Prevalence of the epsilon4 Allele of Apolipoprotein E in Alzheimer's Disease Cohorts. J Mol Neurosci 60(3), 316-324.

Hester, R., Foxe, J.J., Molholm, S., Shpaner, M., Garavan, H., 2005. Neural mechanisms involved in error processing: a comparison of errors made with and without awareness. Neuroimage 27(3), 602-608.

Jiang, J., Liu, T., Crawford, J.D., Kochan, N.A., Brodaty, H., Sachdev, P.S., Wen, W., 2020. Stronger bilateral functional connectivity of the frontoparietal control network in near-centenarians and centenarians without dementia. Neuroimage 215, 116855.
Jin, Y., Lin, L., Xiong, M., Sun, S., Wu, S.C., 2023. Moderating effects of cognitive reserve on the relationship between brain structure and cognitive abilities in middle-aged and older adults. Neurobiol Aging 128, 49-64.

Koshino, H., Osaka, M., Shimokawa, T., Kaneda, M., Taniguchi, S., Minamoto, T., Yaoi, K., Azuma, M., Higo, K., Osaka, N., 2023. Cooperation and competition between the default mode network and frontal parietal network in the elderly. Front Psychol 14, 1140399.

Kuhn, M., 2014. Futility analysis in the cross-validation of machine learning models. arXiv preprint arXiv:1405.6974.

Laird, A.R., Eickhoff, S.B., Li, K., Robin, D.A., Glahn, D.C., Fox, P.T., 2009. Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. J Neurosci 29(46), 14496-14505.

Lindenberger, U., Nagel, I.E., Chicherio, C., Li, S.C., Heekeren, H.R., Backman, L., 2008. Age-related decline in brain resources modulates genetic effects on cognitive functioning. Front Neurosci 2(2), 234-244.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S.G., Dias, A., Fox, N., Gitlin, L.N., Howard, R., Kales, H.C., Kivimaki, M., Larson, E.B., Ogunniyi, A., Orgeta, V., Ritchie, K., Rockwood, K., Sampson, E.L., Samus, Q., Schneider, L.S., Selbaek, G., Teri, L., Mukadam, N., 2020. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 396(10248), 413-446.

Lu, A.T., Hannon, E., Levine, M.E., Crimmins, E.M., Lunnon, K., Mill, J., Geschwind, D.H., Horvath, S., 2017. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. Nat Commun 8, 15353.

Lyall, D.M., Cullen, B., Allerhand, M., Smith, D.J., Mackay, D., Evans, J., Anderson, J., Fawns-Ritchie, C., McIntosh, A.M., Deary, I.J., Pell, J.P., 2016. Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. PLoS One 11(4), e0154222.

Malagurski, B., Liem, F., Oschwald, J., Merillat, S., Jancke, L., 2020. Longitudinal functional brain network reconfiguration in healthy aging. Hum Brain Mapp 41(17), 4829-4845.

Malagurski, B., Liem, F., Oschwald, J., Merillat, S., Jancke, L., 2020. Functional dedifferentiation of associative resting state networks in older adults - A longitudinal study. Neuroimage 214, 116680.

Mather, M., 2021. Noradrenaline in the aging brain: Promoting cognitive reserve or accelerating Alzheimer's disease? Seminars in Cell & Developmental Biology 116, 108-124.

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell, E., Marchini, J., 2021. Computationally efficient whole-genome regression for quantitative and binary traits. Nature Genetics 53(7), 1097-1103.

Mevel, K., Chetelat, G., Eustache, F., Desgranges, B., 2011. The default mode network in healthy aging and Alzheimer's disease. Int J Alzheimers Dis 2011, 535816.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 19(11), 1523-1536.

Minzenberg, M.J., Yoon, J.H., Soosman, S.K., Carter, C.S., 2018. Altered brainstem responses to modafinil in schizophrenia: implications for adjunctive treatment of cognition. Transl Psychiatry 8(1), 58.

Naghavi, H.R., Nyberg, L., 2005. Common fronto-parietal activity in attention, memory, and consciousness: shared demands on integration? Conscious Cogn 14(2), 390-425.

Norton, S., Matthews, F.E., Barnes, D.E., Yaffe, K., Brayne, C., 2014. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol 13(8), 788-794.

Nucci, M., Mapelli, D., Mondini, S., 2012. Cognitive Reserve Index questionnaire (CRIq): a new instrument for measuring cognitive reserve. Aging Clin Exp Res 24(3), 218-226.
Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., et al, 2016. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533(7604), 539-542.

Nyberg, L., Andersson, M., Kauppi, K., Lundquist, A., Persson, J., Pudas, S., Nilsson, L.G., 2014. Age-related and genetic modulation of frontal cortex efficiency. J Cogn Neurosci 26(4), 746-754.

Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A et al., 2016. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533(7604), 539-542.

Opdebeeck, C., Martyr, A., Clare, L., 2016. Cognitive reserve and cognitive function in healthy older people: a meta-analysis. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 23(1), 40-60.

Papenberg, G., Lindenberger, U., Backman, L., 2015. Aging-related magnification of genetic effects on cognitive and brain integrity. Trends Cogn Sci 19(9), 506-514.
Papenberg, G., Salami, A., Persson, J., Lindenberger, U., Bäckman, L., 2015. Genetics and functional imaging: effects of APOE, BDNF, COMT, and KIBRA in aging. Neuropsychol Rev 25(1), 47-62.

Persson, J., Lustig, C., Nelson, J.K., Reuter-Lorenz, P.A., 2007. Age differences in deactivation: a link to cognitive control? J Cogn Neurosci 19(6), 1021-1032.

Plini, E.R.G., O'Hanlon, E., Boyle, R., Sibilia, F., Rikhye, G., Kenney, J., Whelan, R., Melnychuk, M.C., Robertson, I.H., Dockree, P.M., 2021. Examining the Role of the Noradrenergic Locus Coeruleus for Predicting Attention and Brain Maintenance in Healthy Old Age and Disease: An MRI Structural Study for the Alzheimer's Disease Neuroimaging Initiative. Cells 10(7).

Prince, M., Wimo, A.G.M., Ali, G.C., Wu, Y.T., Prina, M., 2015. World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends. London: Alzheimer’s Disease International. pp. 1–87.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3), 559-575.

Rapp, P.R., Amaral, D.G., 1992. Individual differences in the cognitive and neurobiological consequences of normal aging. Trends Neurosci 15(9), 340-345.

Robertson, I.H., 2013. A noradrenergic theory of cognitive reserve: implications for Alzheimer's disease. Neurobiol Aging 34(1), 298-308.

Robertson, I.H., 2014. Right hemisphere role in cognitive reserve. Neurobiol Aging 35(6), 1375-1385.

Robbins, T.W., Arnsten, A.F., 2009. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. Annu Rev Neurosci 32, 267-287.

Shalev, N., Brosnan, M.B., Chechlacz, M., 2020. Right Lateralized Brain Reserve Offsets Age-Related Deficits in Ignoring Distraction. Cereb Cortex Commun 1(1), tgaa049.

Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex 22(1), 158-165.

Singh-Curry, V., Husain, M., 2009. The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. Neuropsychologia 47(6), 1434-1448.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A 106(31), 13040-13045.

Staffaroni, A.M., Brown, J.A., Casaletto, K.B., Elahi, F.M., Deng, J., Neuhaus, J., Cobigo, Y., Mumford, P.S., Walters, S., Saloner, R., Karydas, A., Coppola, G., Rosen, H.J., Miller, B.L., Seeley, W.W., Kramer, J.H., 2018. The Longitudinal Trajectory of Default Mode Network Connectivity in Healthy Older Adults Varies As a Function of Age and Is Associated with Changes in Episodic Memory and Processing Speed. J Neurosci 38(11), 2809-2817.

Stern, Y., 2012. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol 11(11), 1006-1012.
Stern, Y., Alexander, G.E., Prohovnik, I., Mayeux, R., 1992. Inverse relationship between education and parietotemporal perfusion deficit in Alzheimer's disease. Annals of Neurology 32(3), 371-375.

Stern, Y., Arenaza-Urquijo, E.M., Bartres-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W.S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., Vuoksimaa, E., the Reserve, R., Protective Factors, P.I.A.E.D., Conceptual Frameworks, W., 2020. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. Alzheimers Dement 16(9), 1305-1311.

Stuss, D.T., 2011. Functions of the frontal lobes: relation to executive functions. J Int Neuropsychol Soc 17(5), 759-765.

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 Med 12(3), e1001779.

Sundararajan, M., Najmi, A., 2020. The many Shapley values for model explanation. arXiv e-prints, arXiv: 1908.08474.

Suttkus, S., Schumann, A., de la Cruz, F., Bar, K.J., 2021. Working memory in schizophrenia: The role of the locus coeruleus and its relation to functional brain networks. Brain Behav 11(5), e02130.

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58: 267-288.

Tucker-Drob, E.M., Reynolds, C.A., Finkel, D., Pedersen, N.L., 2014. Shared and unique genetic and environmental influences on aging-related changes in multiple cognitive abilities. Dev Psychol 50(1), 152-166.

Valenzuela, M.J., Sachdev, P., 2006. Cognitive leisure activities, but not watching TV, for future brain benefits. Neurology 67(4), 729; author reply 729.

Valenzuela, M.J., Sachdev, P., 2006. Brain reserve and dementia: a systematic review. Psychol Med 36(4), 441-454.
VanderWeele, T.J., 2016. Mediation Analysis: A Practitioner's Guide. Annual Review of Public Health 37(1), 17-32.

van Donkelaar, M.M.J., Hoogman, M., Shumskaya, E., Buitelaar, J.K., Bralten, J., Franke, B., 2020. Monoamine and neuroendocrine gene-sets associate with frustration-based aggression in a gender-specific manner. Eur Neuropsychopharmacol 30, 75-86.

van Loenhoud, A.C., Wink, A.M., Groot, C., Verfaillie, S.C.J., Twisk, J., Barkhof, F., van Berckel, B., Scheltens, P., van der Flier, W.M., Ossenkoppele, R., 2017. A neuroimaging approach to capture cognitive reserve: Application to Alzheimer's disease. Hum Brain Mapp 38(9), 4703-4715

Varela-Lopez, B., Cruz-Gomez, A.J., Lojo-Seoane, C., Diaz, F., Pereiro, A.X., Zurron, M., Lindin, M., Galdo-Alvarez, S., 2022. Cognitive reserve, neurocognitive performance, and high-order resting-state networks in cognitively unimpaired aging. Neurobiol Aging 117, 151-164.

Veldsman, M., Tai, X.Y., Nichols, T., Smith, S., Peixoto, J., Manohar, S., Husain, M., 2020. Cerebrovascular risk factors impact frontoparietal network integrity and executive function in healthy ageing. Nat Commun 11(1), 4340**.**

Vidal-Pineiro, D., Valls-Pedret, C., Fernandez-Cabello, S., Arenaza-Urquijo, E.M., Sala-Llonch, R., Solana, E., Bargallo, N., Junque, C., Ros, E., Bartres-Faz, D., 2014. Decreased Default Mode Network connectivity correlates with age-associated structural and cognitive changes. Front Aging Neurosci 6, 256.

Wedemeyer, C., Goutman, J.D., Avale, M.E., Franchini, L.F., Rubinstein, M., Calvo, D.J., 2007. Functional activation by central monoamines of human dopamine D(4) receptor polymorphic variants coupled to GIRK channels in Xenopus oocytes. Eur J Pharmacol 562(3), 165-173.

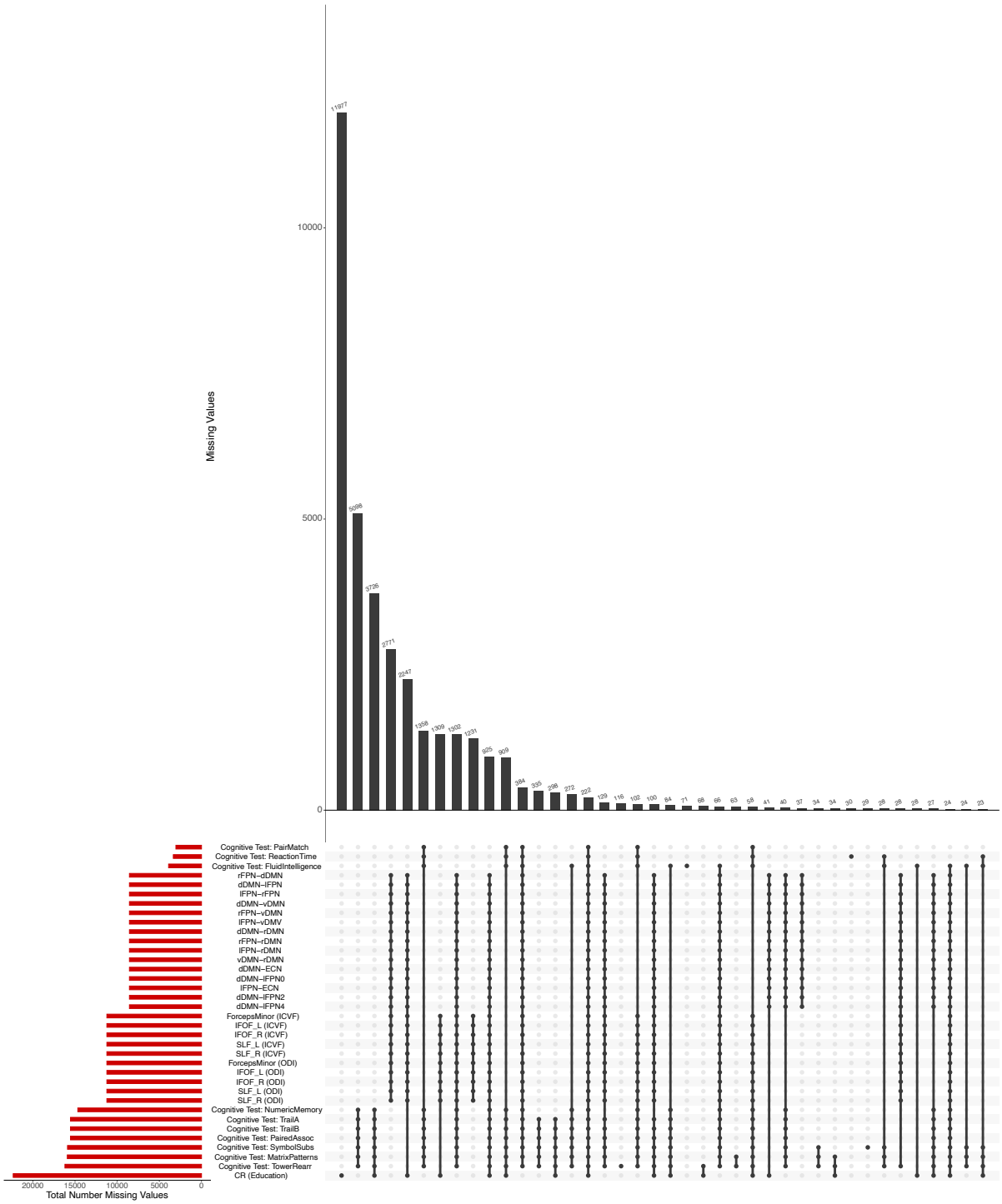
Zhukovsky, P., Coughlan, G., Buckley, R., Grady, C., Voineskos, A.N., 2023. Connectivity between default mode and frontoparietal networks mediates the association between global amyloid-beta and episodic memory. Hum Brain Mapp 44(3), 1147-1157.

SUPPLEMENTARY MATERIALS

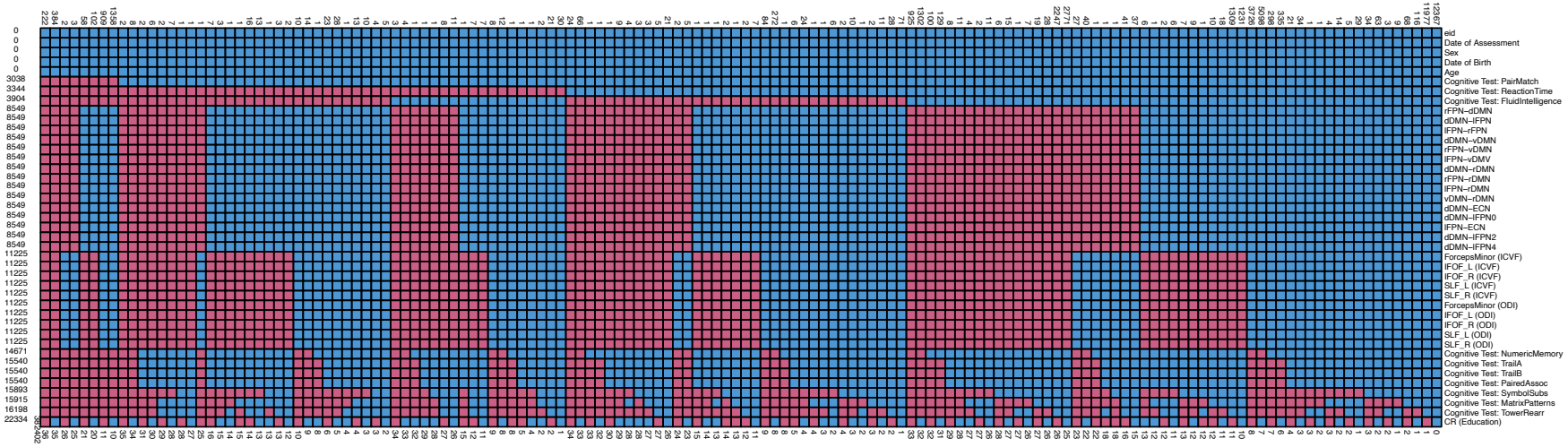
Supplementary Table 1.

Features included in the analysis before preprocessing as separated into the datasets summarized in Figure 1 (Demographics with AgeXSex, imaging derived penotypes representing structure and functional front-parietal connectivity (FPC IDPs), cognitive tests and education information)

				Missing education information	
Variables	Total N (complete cases %)	Levels	Total	No	Yes
Total N (%)			48561	26227 (54.0)	22334 (46.0)
Demographics					
Sex	48561 (100.0)	Female	25078 (51.6)	13969 (53.3)	11109 (49.7)
		Male	23483 (48.4)	12258 (46.7)	11225 (50.3)
Age	48561 (100.0)	Mean (SD)	64.7 (7.7)	65.1 (7.7)	64.2 (7.7)
Cognitive Tests					
ReactionTime	45217 (93.1)	Mean (SD)	597.1 (110.7)	605.1 (113.4)	587.8 (106.8)
NumericMemory	33890 (69.8)	Mean (SD)	6.7 (1.5)	6.4 (1.6)	6.9 (1.4)
FluidIntelligence	44657 (92.0)	Mean (SD)	6.6 (2.1)	6.0 (1.9)	7.2 (2.0)
TrailA	33021 (68.0)	Mean (SD)	226.3 (90.5)	231.8 (98.2)	220.4 (80.9)
TrailB	33021 (68.0)	Mean (SD)	558.9 (284.0)	591.2 (313.8)	524.2 (243.3)
MatrixPatterns	32646 (67.2)	Mean (SD)	7.9 (2.1)	7.4 (2.1)	8.5 (2.0)
TowerRearr	32363 (66.6)	Mean (SD)	9.8 (3.2)	9.5 (3.3)	10.2 (3.1)
SymbolSubs	32668 (67.3)	Mean (SD)	18.7 (5.3)	18.0 (5.4)	19.5 (5.2)
PairedAssoc	33021 (68.0)	Mean (SD)	6.8 (2.6)	6.4 (2.7)	7.4 (2.5)
PairMatch	45523 (93.7)	Mean (SD)	3.7 (3.0)	3.7 (3.0)	3.6 (3.0)
FPC IDPs					
ForcepsMinor (ICVF)	37336 (76.9)	Mean (SD)	0.5 (0.0)	0.5 (0.0)	0.5 (0.0)
IFOF_L (ICVF)	37336 (76.9)	Mean (SD)	0.5 (0.0)	0.5 (0.0)	0.5 (0.0)
IFOF_R (ICVF)	37336 (76.9)	Mean (SD)	0.5 (0.0)	0.5 (0.0)	0.5 (0.0)
SLF_L (ICVF)	37336 (76.9)	Mean (SD)	0.6 (0.0)	0.6 (0.0)	0.6 (0.0)
SLF_R (ICVF)	37336 (76.9)	Mean (SD)	0.6 (0.0)	0.6 (0.0)	0.6 (0.0)
ForcepsMinor (ODI)	37336 (76.9)	Mean (SD)	0.2 (0.0)	0.2 (0.0)	0.2 (0.0)
IFOF_L (ODI)	37336 (76.9)	Mean (SD)	0.2 (0.0)	0.2 (0.0)	0.2 (0.0)
IFOF_R (ODI)	37336 (76.9)	Mean (SD)	0.2 (0.0)	0.2 (0.0)	0.2 (0.0)
SLF_L (ODI)	37336 (76.9)	Mean (SD)	0.2 (0.0)	0.2 (0.0)	0.2 (0.0)
SLF_R (ODI)	37336 (76.9)	Mean (SD)	0.2 (0.0)	0.2 (0.0)	0.2 (0.0)
dDMN-rFPN	40012 (82.4)	Mean (SD)	-0.8 (1.0)	-0.7 (1.0)	-0.8 (1.0)
dDMN-lFPN	40012 (82.4)	Mean (SD)	-0.5 (0.9)	-0.5 (0.9)	-0.5 (0.9)
rFPN-lFPN	40012 (82.4)	Mean (SD)	1.3 (1.0)	1.3 (1.0)	1.2 (1.0)
dDMN-vDMN	40012 (82.4)	Mean (SD)	1.6 (0.9)	1.6 (0.9)	1.5 (0.9)
rFPN-vDMN	40012 (82.4)	Mean (SD)	0.3 (0.9)	0.3 (0.9)	0.3 (0.9)
lFPN-vDMV	40012 (82.4)	Mean (SD)	-0.1 (0.9)	-0.2 (0.9)	-0.1 (0.9)
dDMN-rDMN	40012 (82.4)	Mean (SD)	2.5 (0.9)	2.5 (0.9)	2.5 (0.9)
rFPN-rDMN	40012 (82.4)	Mean (SD)	0.3 (1.0)	0.4 (1.0)	0.3 (1.0)
lFPN-rDMN	40012 (82.4)	Mean (SD)	1.1 (0.9)	1.1 (0.9)	1.1 (0.9)
vDMN-rDMN	40012 (82.4)	Mean (SD)	-0.8 (0.9)	-0.8 (0.9)	-0.8 (0.9)
dDMN-ECN	40012 (82.4)	Mean (SD)	-1.4 (1.0)	-1.4 (1.0)	-1.5 (1.0)
rFPN-ECN	40012 (82.4)	Mean (SD)	0.2 (0.9)	0.2 (0.9)	0.2 (1.0)
lFPN-ECN	40012 (82.4)	Mean (SD)	-1.2 (1.0)	-1.2 (1.0)	-1.2 (1.0)
vDMN-ECN	40012 (82.4)	Mean (SD)	-0.0 (1.0)	-0.0 (0.9)	-0.0 (1.0)
rDMN-ECN	40012 (82.4)	Mean (SD)	1.4 (0.9)	1.4 (0.9)	1.3 (0.9)



Supplementary Figure 1A: Missing values and missing at random analysis.
Missing values pattern exploration using R package *UpSetR* showing the combinations of missingness across cases

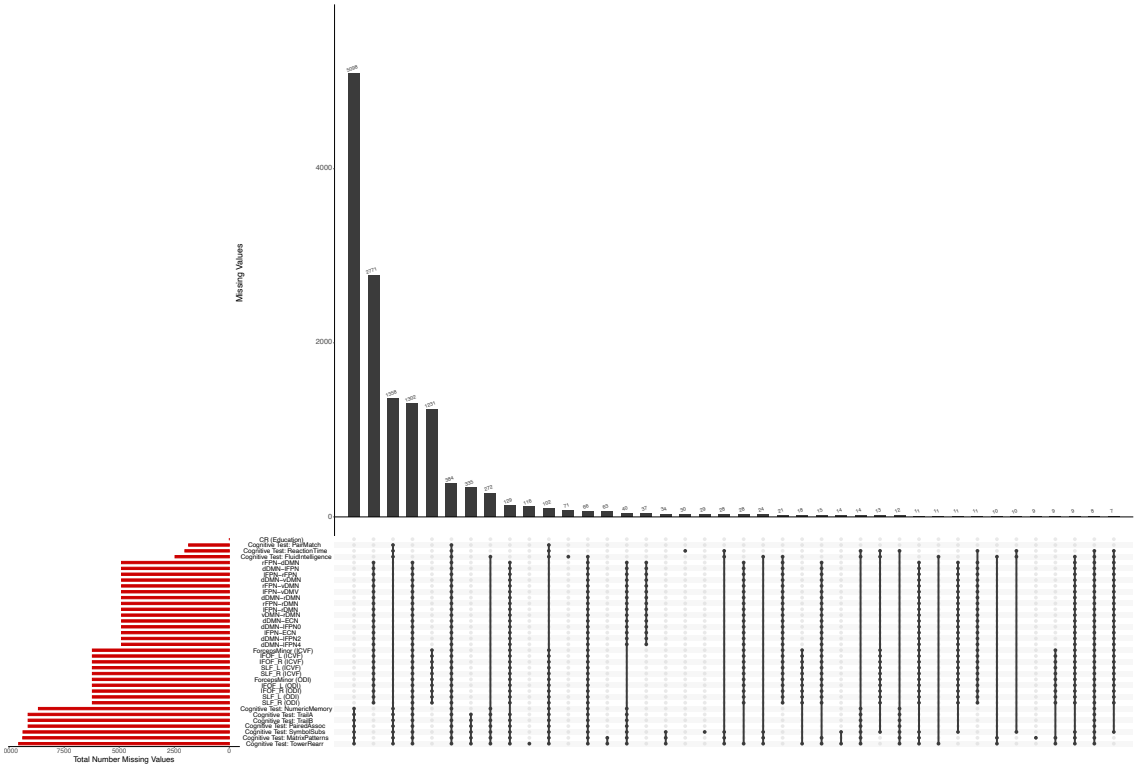


Supplementary Figure 1B: Missing values and missing at random analysis

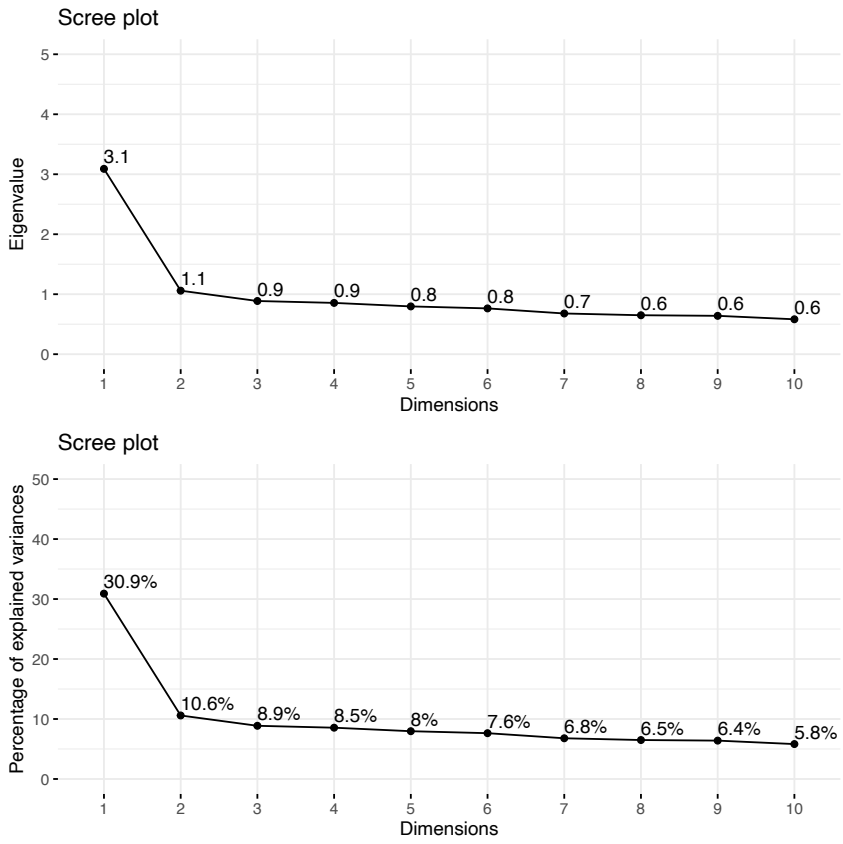
Missing values pattern exploration using R package *mice* and *finalfit*. Missing values in red and complete in blue, with total numbers of participants found to the left and total missing features (red squares) to the right. Total missing values per feature found at the bottom.

A.

B.



Supplementary Figure 2: Missing values and missing at random analysis.
Missing values pattern exploration using R package *UpSetR* and *naniar*, showing the combinations of missingness across cases. Maximimum of missing cases per column is the cognitive test of tower rearrangement with (36.47%) missing data

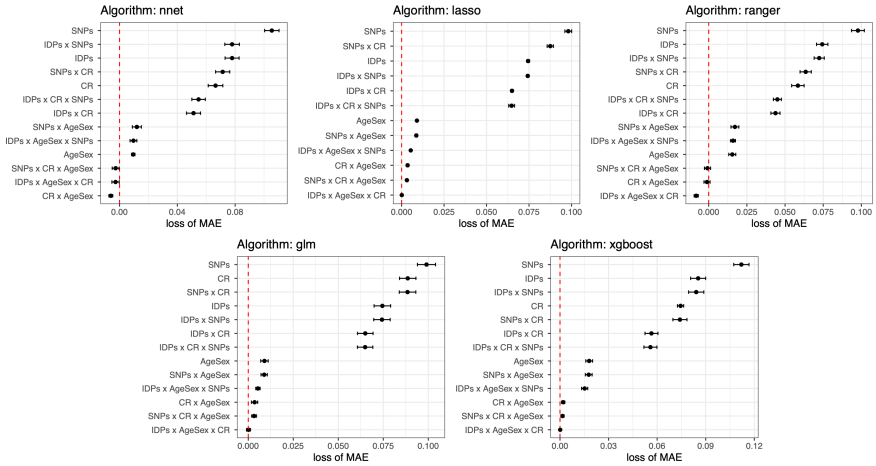


Supplementary Figure 3. PCA scree plots. See Methods section for further information.

Supplementary Table 2 PCA coordinate s(loading* standard deviation) using *factoextra*

	PC 1	PC 2	PC 3
ReactionTime	-0.384	0.496	-0.473
NumericMemory	0.512	0.429	0.084
FluidIntelligence	0.636	0.412	0.018
TrailA	-0.544	0.379	-0.069
TrailB	-0.587	0.17	-0.076
MatrixPatterns	0.646	0.185	0.01
TowerRearr	0.626	0.002	-0.077
SymbolSubs	0.652	-0.231	0.056
PairedAssoc	0.495	0.33	0.006
PairMatch	-0.395	0.3	0.798

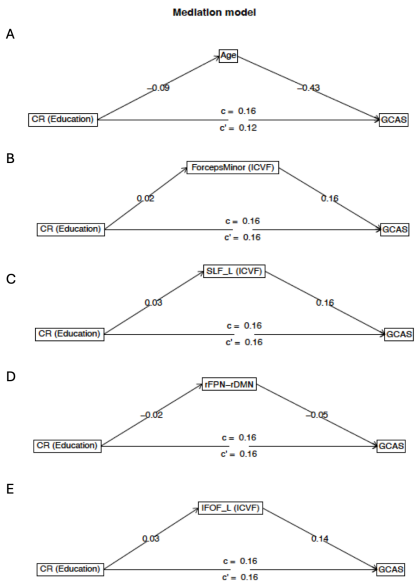
A



B

Mediator	Effect of CR on mediator (a)	Effect of mediator (b)	Indirect effect (ab)	95% CI	
				Lower	Upper
rs75618936(SYT1)	0.0022 (8e-04)	0.0065 (0.0454)	0 (1e-04)	-2e-04	2e-04
rs78005227(SYT1)	0.0025 (8e-04)	-0.0408 (0.0443)	-1e-04 (1e-04)	-4e-04	1e-04
rs79104808(SYT1)	0.0022 (8e-04)	0.0023 (0.0452)	0 (1e-04)	-2e-04	2e-04
Age	-0.0949 (0.0061)	-0.4319 (0.0055)	0.0411 (0.0029)	0.0356	0.0468
rs74560629(SYT1)	0.0022 (8e-04)	0.0023 (0.0452)	0 (1e-04)	-2e-04	2e-04
rs62465540	-1e-04 (2e-04)	0.136 (0.1693)	0 (0)	-1e-04	1e-04
rs62468660	-1e-04 (2e-04)	0.1117 (0.1745)	0 (0)	-1e-04	1e-04
rs10224002(PRKAG2)	-0.0014 (0.0016)	0.0162 (0.0233)	0 (1e-04)	-2e-04	1e-04
ForcepsMinor (ICVF)	0.0235 (0.0062)	0.1576 (0.006)	0.0037 (0.001)	0.0017	0.0057
SLF_L (ICVF)	0.0305 (0.0062)	0.1602 (0.006)	0.0049 (0.001)	0.0029	0.007
rFPN-rDMN	-0.0239 (0.0062)	-0.0466 (0.0061)	0.0011 (3e-04)	5e-04	0.0018
rs10254101(PRKAG2)	0.0032 (0.0029)	0.0405 (0.0128)	1e-04 (1e-04)	-1e-04	4e-04
rs8056290(PRKCB)	-0.0029 (0.0031)	-0.0483 (0.0123)	1e-04 (2e-04)	-1e-04	5e-04
dDMN-ECN	-0.0026 (0.0062)	0.0562 (0.0061)	-2e-04 (3e-04)	-8e-04	5e-04
rs3785383(PRKCB)	0.0027 (0.0024)	0.0225 (0.0159)	1e-04 (1e-04)	-1e-04	3e-04
IFOF_L (ICVF)	0.032 (0.0062)	0.1442 (0.006)	0.0046 (9e-04)	0.0029	0.0065
Sex	0.0071 (0.0031)	-0.026 (0.0122)	-2e-04 (1e-04)	-5e-04	0
rs2888788(PRKAG2)	-0.0068 (0.0031)	-0.0557 (0.0122)	4e-04 (2e-04)	0	8e-04
rs7786782(PRKAG2)	-0.0011 (0.003)	0.0438 (0.0126)	0 (1e-04)	-3e-04	2e-04
rs6947064(PRKAG2)	9e-04 (0.002)	0.043 (0.0188)	0 (1e-04)	-1e-04	2e-04
rs80299807	-0.0062 (0.0031)	-0.057 (0.0122)	4e-04 (2e-04)	0	8e-04
rs1872691(BRD7-ADCY7)	0.0033 (0.0027)	0.0087 (0.0138)	0 (1e-04)	-1e-04	2e-04

C



Supplementary Figure 4. Imputed Dataset (K-nearest neighbours using 3 neighbours in tidymodels) Given that only 20.2% of our dataset is missing but 54% (14,151/26,227) of participants are deleted from our analysis in complete case analysis, we performed a sensistivity analysis with imputed data to understand how and if results differed substantially with more data. As seen below,same analaysis was performed (Figure 1) and the combination of IDPs, AgeSex and CR came as best performing too in most algorithms as is the case for the complete case analysis (Figure 2). Feature importance was assessed in the same way with top performing features seen in B. 17 features differed (8 SNPs only selected in imputed data analysis and 9 IDPs and SNPs only found in complet-case analysis, with 14 chosen in common). Regarding mediation analysis, rFPN-rDMN and age were found as significant in both ascertaining their robustness. Mediation of all significant features just in imputation data analysis can be seen here in C.

Supplementary Figure 5: Continous and categorical distribution of selected features with information on cognitive tests, demographics, education, genes and multimodal neuroimaging data.

Appendix

Fields used

Demographic information and cognitive data

f.34.0.0:Clinical_Birth_Year
f.52.0.0:Clinical_Birth_Month
f.53.2.0:Clinical_Date_Assessment
f.31.0.0:Clinical_Sex
f.845.0.0:Education_0
f.845.1.0:Education_1
f.845.2.0:Education_2
f.20023.2.0:Cog_ReactionTime
f.4282.2.0:Cog_NumericMemory
f.20016.2.0:Cog_FluidIntelligence
f.6348.2.0:Cog_TrailA
f.6350.2.0:Cog_TrailB
f.6373.2.0:Cog_MatrixPatterns
f.21004.2.0:Cog_TowerRearr
f.23324.2.0:Cog_SymbolSubs
f.20197.2.0:Cog_PairedAssoc
f.399.2.2:Cog_PairMatch

MRI data

f.25661.2.0:WM_ICVF_ForcepsMinor
f.25662.2.0:WM_ICVF_IFOF_Left
f.25663.2.0:WM_ICVF_IFOF_Right
f.25671.2.0:WM_ICVF_SLF_Left
f.25672.2.0:WM_ICVF_SLF_Right
f.25688.2.0:WM_OD_ForcepsMinor
f.25689.2.0:WM_OD_IFOF_Left
f.25690.2.0:WM_OD_IFOF_Right
f.25698.2.0:WM_OD_SLF_Left
f.25699.2.0:WM_OD_SLF_RightSupplementary

The 25 components group-ICA spatial maps and connection edges are available online (http://www.fmrib.ox.ac.uk/ukbiobank).

C7, c11, c15, c16, c20, c21, c29, c33, c34, c35, c106, c110, c111, c112,c114