Thirteen independent genetic loci associated with preserved processing speed in a study of cognitive resilience in 330,097 individuals in the UK Biobank

Joan Fitzgerald[1], Laura Fahey[1,2], Laurena Holleran[1], Pilib Ó Broin[2], Gary Donohoe[1], Derek W. Morris[1]

[1] Cognitive Genetics and Cognitive Therapy Group, Centre for Neuroimaging & Cognitive Genomics, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland.

[2] School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, Ireland.

Corresponding author:

Dr Derek Morris, Room 106, Discipline of Biochemistry, National University of Ireland Galway,

University Road, Galway, Ireland.

Tel: + 353 91 494439; Email: derek.morris@nuigalway.ie

Abstract

Cognitive resilience is the ability to withstand the negative effects of stress on cognitive functioning and is important for maintaining quality of life while aging. UK Biobank (UKB) does not have direct measurements of the same cognitive phenotype at distal timepoints. Therefore, we used number of education years (EY) as a proxy phenotype for past cognitive performance. Current cognitive performance was determined based on processing speed. This approach captured an average time span of 40 years between past and current cognitive performance in 330,097 individuals. A confounding factor was that EY is highly polygenic and masked the genetics of resilience. To overcome this, we employed Genomics Structural Equation Modelling (GenomicSEM) to perform a GWAS-by-subtraction using two GWAS, one GWAS of EY and resilience and a second GWAS of EY but not resilience. Subtracting one from the other generated a GWAS of *Resilience*. Replication of this approach was shown using independent discovery and replication samples within UKB and the full UKB GWAS results show significant genetic correlation with a GWAS of cognitive change in the independent Health and Retirement Study (N=9,526; P=1.5x10^{-3}). We found 13 independent genetic loci for *Resilience*. Functional analyses showed enrichment in several brain regions and involvement of specific cell types, including GABAergic neurons (P=6.59x10^{-8}) and glutamatergic neurons (P=6.98x10^{-6}) in the cortex. Gene-set analyses implicated the biological process “neuron differentiation” (P=9.7x10^{-7}) and the cellular component “synaptic part” (P=2.14x10^{-6}). Mendelian randomization analysis showed a causative effect of white matter volume on cognitive resilience. These results enhance neurobiological understanding of resilience.

Introduction

Cognitive decline is one of the most feared aspects of aging leading to major health and social issues and is associated with illness, dementia and death [1]. Non-pathological or age-related cognitive decline leads to increased challenges in completing tasks that require information processing and memory, which in turn leads to a deleterious effect on an individual's enjoyment of and participation in life events [2] . Cognitive resilience is our ability to withstand negative effects of stress and maintain cognitive functioning. Understanding the factors that contribute to resilience is becoming increasingly important given the aging demographics of the world's population [3]. There is a growing knowledge of how non-genetic factors such as cardiovascular health and social participation contribute to cognitive resilience [4]; however an understanding of the genetic contribution has been hampered by the lack of large datasets with genetic data and suitable longitudinal data on cognition. One theory that examines the biological influences on rates of cognitive decline in healthy aging is the concept of reserve, maintenance, and compensation leading to cognitive resilience[5]. Reserve is usually described in terms of both brain reserve, which is the overall strength of size of structural components such as quantity of neurons and synapses and cognitive reserve, which refers to adaptability of these components [6]. These are hypothesised to reflect a level of neural resources built up over our lifetime, maintained via the ability to repair cellular damage to maintain cognitive function, with losses compensated for by use of alternative undamaged cognitive functions. In turn these mechanisms are thought to be mediated by a combination of environmental and genetic factors.

Others propose that variation in the rate of cognitive decline can be explained by variation in intelligence. Longitudinal analysis in the Lothian Birth Cohort has shown that childhood intelligence has a protective effect on cognitive decline in late life [7]. Other studies show that while higher education reflects greater cognitive ability, the rates of change in that ability over time are consistent across all education levels, with those starting at a higher level simply having further to fall before they present with mild cognitive impairment [8] [9]. The role of intelligence is confounded by the fact that higher intelligence is associated with healthier life styles, which has a protective effect on cognitive decline [10].

Salthouse proposed the reduced speed of processing hypothesis as earlier studies on cognitive decline showed that processing speed is one of the strongest predictors of performance across cognitive tasks in older adults [11, 12]. This theory proposes that older adults take longer to process information and the result of this slower processing leads to impairment in cognitive

functions and information is not available for the next part of a task as quickly as in younger adults. It is proposed that superior intelligence is linked to faster processing speed and speed of higher order information processing explains about 80% of variance in cognitive ability [13]. In a study using 1,800 adults ranging in age from 20 to 90, it was found that 70 to 80% of decline in processing speed was shared with declining reasoning ability [14].

The purpose of this study was to explore genetic variation associated with cognitive resilience within the UK Biobank (UKB)[15]. We first used reaction time (RT), reflecting an individual's processing speed, as a measure of current cognitive performance. Processing speed is a key component, and predictor, of cognitive ability [13] [16]. In the absence of a direct measure of processing speed at an earlier timepoint, we used academic achievement measured by number of years in education (education years (EY)) as a proxy phenotype for cognitive performance in early adulthood, following several previous studies [17-19]. Individual differences in processing speed are important in the relationship between executive functioning and academic performance [20]. This approach captures an average time span of 40 years between past and current cognitive performance in UKB. A confounding factor in this strategy is that EY is highly heritability with a polygenic nature [21] that can mask the genetics of resilience. To overcome this we employed Genomics Structural Equation Modelling (GenomicSEM) [22] to perform a GWAS-by-subtraction [23] using two GWAS, one which captured genetic variants associated with EY and resilience and a second which captured genetic variants associated with EY but not resilience. Subtracting one from the other generated two new GWAS, one capturing EY and the other capturing the genetics of a processing speed-based cognitive resilience phenotype. Replication of this approach was shown using independent discovery and replication samples within UKB and explored in an independent longitudinal data set. Full GWAS results were examined further using functional genomics analysis.

Results

Initial phenotype development

Figure 1 shows an overview of the analysis steps and a detailed description of the process used to generate the resilience variable is included in the online methods. Given the multi-step method proposed in this analysis, we sought to confirm findings using our method in an independent sample. Therefore, we divided the UKB into discovery (n=266,543; 81% of participants) and replication (n=63,554; 19% of participants) samples (Figure 1a). Sample sizes used for analysis are shown in Supplementary Table 1. We used EY as a proxy phenotype measuring past cognitive performance [13]. EY was calculated using a combination of data supplied on years in education and educational attainment (see online methods). Processing speed as measured by RT was chosen as an indicator of current performance given its strong correlation with age and the fact that data was available on most participants in UKB. We created a binary variable for each measure by using the average score within the dataset to split the participants into similarly sized groups. EY was split into above and below average based on participants completing greater than or equal to 17 years, or less than 17 years in education. RT was corrected for age and normalised using a log 10 transformation and was split into faster and slower based on participants having a processing speed better or worse than the mean value (see online methods), such that faster RT speeds reflected better processing speed, and thus cognitive performance, than slower RT. By combining these two binary variables, we created four groups of participants (Figure 1b). One of these groups demonstrated high resilience and these were our cases for GWAS who had below average EY previously and faster than average RT now. A second group demonstrated low resilience or cognitive decline, and these were our controls for GWAS who had above average EY previously and slower than average RT now. Results for this GWAS were dominated by SNPs associated with EY because the high resilience cases and low resilience controls had below average and above average EY measures, respectively. We named this GWAS "EY+Res" because it identified SNPs associated with both EY and resilience (Figure 1c).

In order to identify those SNPs that were associated with resilience alone and remove those SNPs that were associated with the EY component of the phenotype, we performed a second GWAS using the two remaining groups of UKB samples that displayed consistent (i.e., unchanging) performance over time. The first of these groups consisted of those with below

average EY previously and slower than average RT now (i.e., consistently below average performance over time); the second group consisted of those who showed above average EY previously and faster than average RT now (i.e., consistently above average performance over time). We named this GWAS "EY/NonRes" because it identified SNPs associated with EY but not resilience (Figure 1d).

We then used GWAS-by-subtraction (GBS) [23] to subtract the results of EY/NonRes from EY+Res to leave SNP associations with resilience. This method uses GenomicSEM [22] to integrate GWAS into structural equation modelling. Following the process described by Demange et al [23], we defined a Cholesky model (Figure 2) using the summary statistics from the EY+Res and EY/NonRes GWASs. Both EY+Res and EY/NonRes were regressed on a latent factor, which captured the shared genetic variance in EY (hereafter "*EduYears*"). EY+Res was further regressed on a second latent factor capturing the variance in EY+Res independent of EY/NonRes, hereafter "*Resilience*". Genetic variance in *Resilience* was independent of genetic variance in *EduYears* (r_g=0) as the *Resilience* factor represents residual genetic variation in our EY+Res phenotype that is not accounted for by the *EduYears* factor. These two latent variables, *Resilience* and *EduYears* were then regressed on each SNP in the original GWASs (EY+Res and EY/NonRes) resulting in new GWAS summary statistics for both *Resilience* and *EduYears* (Figure 1e).

Discovery and replication analysis within UKB

For the discovery sample, we performed the two initial GWASs (discovery.EY+Res and discovery.EY/NonRes) and then performed GBS on both sets of samples resulting in *discovery.Resilience* GWAS results and *discovery.EduYears* GWAS results. We repeated this for the replication sample to produce *replication.Resilience* GWAS results and *replication.EduYears* GWAS results. Comparison of the *discovery.Resilience* GWAS with the *replication.Resilience* GWAS by LD score regression (LDSR) analysis [22] showed extremely high correlation between the two data sets (r_g = 0.964, P = 4.45 x 10^{-44}). The *discovery.Resilience* GWAS was then processed through FUMA v 1.3.6 [24] and ten independent genome-wide significant SNPs were identified. When compared to the *replication.Resilience* GWAS, there was a consistent direction of effect for all ten SNPs (Binomial sign test, P = 9.77 x 10^{-4}). Five of the ten SNPs were significant after Bonferroni multiple test correction for those SNPs tested (P < 0.005). Thus, we demonstrated that we could replicate genetic associations with *Resilience* in an independent sample. Results for the

ten-independent genome-wide significant SNPs and their replication analysis are in Supplementary Table 2.

Analysis of the full sample

Next, we combined both the discovery and replication samples to run an analysis on the full sample (n=330,097). This resulted in initial EY+Res and EY/NonRes GWASs and following GBS, *Resilience* GWAS results and *EduYears* GWAS results. SNP based heritability estimate analysis showed a h^2 value of 0.13 (SE = .006) for *Resilience*. For comparison in similarly sized samples, we also ran GWASs of EY and RT using participants randomly selected from UKB (EY, n=82,000 above average EY cases and n=81,999 below average EY controls; RT, n=82,000 faster than average RT cases and n=82,000 slower than average RT controls). These comparisons are shown in Supplementary Table 3. A Manhattan plot and a Quantile-quantile (Q-Q) plot of *Resilience* on the full sample is shown in Figure 3a and 3b. Manhattan plots for the other five GWAS (EY+Res, EY/NonRes, *EduYears*, EY and RT) are in Supplementary Figure 1.

Initially, both EY+Res and EY/NonRes had a strong negative correlation with EY (r_g = -0.88 and r_g = -0.89 respectively (Supplementary Figure 2). The strength of these correlations likely reflects the major contribution of EY to these phenotypes and they are negative because for EY+Res and EY/NonRes, the direction of effect is in the opposite direction to EY, as the cases are low EY whereas for the EY GWAS, the cases are high EY. EY+Res and EY/NonRes had a moderate positive correlation with each other (r_g = 0.54). After GBS there was no genetic correlation between *Resilience* and *EduYears* (r_g = 0.01, P = 0.803) suggesting that the subtraction had successfully separated out the genetic associations for both phenotypes.

Although the EY component of *Resilience* was addressed by the GBS method, the RT component was not and the genetic correlation between *Resilience* and RT was strong (r_g = 0.80; Supplementary Figure 2). This finding was examined further following functional analysis of associated loci with detail on this provided at the end of Results.

Analysis in a dataset independent of UKB

The US Health and Retirement study (HRS) is a longitudinal study of adults aged 50 years or older in households in the United States [25]. The study commenced in 1992 and participants were interviewed at baseline and every two subsequent years (see online methods). In order

to maximise sample size, we used immediate memory recall as our cogntive measure. Using linear effect modelling [26] we generated a cognitive change phenotype and ran a GWAS using available participants (N=9,526). Although no significant genetic loci were found in this analysis and only one of our *Resilience* genetic loci was nominally significant in the HRS GWAS (chromosome 3 locus; P = 0.025), there was a significant genetic correlation (r_g = -.65, P = 1.5×10^{-3}) between cogntive change and *Resilience*. This negative correlation is expected as the HRS GWAS used a measure of cognitive change whereas the *Resilience* GWAS used a measure of resilience to change. Manhattan and Q-Q plots of the HRS cognitive change GWAS are in Supplementary Figure 1g.

Functional analysis

Description of genetic loci

Function analysis was performed on *Resilience* in FUMA v 1.3.6 [24]. (Note: see Online Methods for parameters used and results publicly available in FUMA ID:171).

A total of 1,329 significant SNPs were tagged from the *Resilience* GWAS and were associated with 26 independent lead SNPs ($P < 5 \times 10^{-8}$). Including SNPs in the reference panel that are in LD with the independent SNPs, resulted in a total of 1,922 candidate SNPs. Functional annotation of the candidate SNPs showed that 82% were intergenic/intronic. A total of 84 SNPs had a Combined Annotation Dependent Depletion (CADD) score greater than the threshold of 12.37 which indicates that the variants are potentially pathogenic [27] (see Supplementary Table 4 and 5).

Lead SNPs were grouped into 13 independent genetic loci that are on 9 different chromosomes. Detailed maps of each locus are available in Supplementary Figure 3. Conditional analyses showed that the significance of all independent lead SNPs at each locus was reduced when the GWAS was conditioned for the index or most associated SNP, confirming the linkage of the index SNP to each lead SNP (Supplementary Table 6).

Fine Mapping

FINEMAP [28] was used to provide further information on significant SNPs in LD with the index SNP on each locus using the GWAS SNPs generated by FUMA (Supplementary Table 7). The $\log_{10}$ Bayes factor (B_{10}) quantifies causal evidence for a particular SNP and a posterior probability value yielding a B_{10} greater than 2 indicates considerable evidence of causality [28]. One SNP, rs62074125, on chromosome 17, exceeded this value (B_{10} = 2.64).

This SNP is an intron within the *WNT3* gene, which is associated with cognitive function [29]. The next highest result was on chromosome 4 where rs2189234 had a value slightly below 2 (B_{10} = 1.62). This SNP is an intronic variant in the *TET2* gene, which is discussed below. FINEMAP analysis showed that the index SNP had the highest Bayes Factor for all loci with four exceptions (Supplementary Table 8).

Gene mapping

Three approaches were used in FUMA to map the associated variants to genes: (a) Positional mapping mapped 141 SNPs to genes based on their genomic location within a 10 kilobase window of known gene boundaries. (b) Expression quantitative trait (eQTL) analysis mapped 207 cis-eQTL SNPs to genes whose expression they were associated with. (c) Chromatin interaction analysis using the 3D DNA to DNA interactions mapped SNPs to 243 genes. Circos plots for all loci are included in Supplementary Figure 4. The circos plot from chromosome 3 shows that 102 genes were mapped to this region, representing 42% of the total genes mapped. In addition, the circos plot from chromosome 17 shows two distinct clusters of SNPs. Genes in this region (*MAPT, WNT3, CRHR1, KANSL1,* and *NSF*) have been previously associated with general cognitive function but also with other cognitive indicators [29]. Details of this gene mapping analysis is in Supplementary Table 9.

In addition to the three approaches above we also performed a genome-wide gene-based association analysis (GWGAS) using the MAGMA function within FUMA [24], which looks at the aggregate association results of all SNPs in a gene in contrast to the previous analyses that examined the association signals at the level of individual SNPs. A GWGAS was performed using the *Resilience* GWAS on 18,879 protein-coding genes containing at least one SNP from the GWAS. Based on the number of genes tested, a Bonferroni-corrected threshold of $P < 2.65 \times 10^{-6}$ was used (see Q-Q plot of this association – Figure 3c). A total of 52 protein coding genes were identified as associated, 40 of which were identified by the previously described strategies (Supplementary Table 10). In total, 33 genes were identified by all four mapping strategies (Figure 3c and Supplementary Table 11).

Many of these 33 genes have been connected with cognitive performance, neurodegenerative disorders or aging and represent potential therapeutic targets: *STAU1* (chr 20) and *SEMA3F* (chr 3) are predicted to control cognitive decline in aging through formation of neural circuits and synaptic transmission [30]. *BNS* (chr 3) codes for bassoon presynaptic cytomatrix protein which is implicated in the regulation of neurotransmitters at inhibitory and excitatory

synapses [31]. *IP6K1* (chr 3) codes for inositol pyrophosphate biosynthesis, and mouse studies have shown its involvement in short term memory by altering presynaptic vesicle release and short-term facilitation of glutamatergic synapses in the hippocampus [32]. *MST1* (chr 3) has been shown to play a role in protecting cells from oxidative stress which leads to aging and eventual cell death [33]. *TET2* (chr 4) codes for ten eleven translocation methyl cytosine dioxygenase 2 which catalyses the production of 5-hydroxymethylcytosine and is associated with increased neurogenesis in the hippocampus and cognition in animal studies [34]. *ATXN2* (chr 20) is involved in regulating mRNA and is linked to decline in cognitive function in older adults [35], general cognitive function [29] and neurodegenerative disorders [36]. The *ATXN2/BRAP* locus has a strong association with parenteral lifespan [37]. Another mapped gene close to *ATXN2* and *BRAP* is *SH2B3*, which encodes lymphocyte adaptor protein LNK, and plays a role in human aging though the mechanism involved is not fully understood [38]. The gene *ALDH2* (chr 12) codes for aldehyde dehydrogenase and there is a link between this enzyme and life span as well as cardiovascular aging [39].

Among the associated SNPs at the 33 prioritized genes are two UTR3 variants on chromosome 3 (rs2681781 (CADD=17.77) and rs4625 (CADD=15.6)) that map to *RBM5* and *DAG1* respectively. Animal studies have shown that *RBM5* is a likely regulator of Rab4a, which in involved in many neurobiological functions including the transport of transmembrane proteins required for neurotransmission [40]. *DAG1* has been associated with increased cognitive performance and is associated with GABAergic signalling in the hippocampus [41]. In addition, one other variant of note is rs1130146 that maps to *DDX27* (chr 20), a gene that was mapped by all strategies except for GWGAS and is associated with longevity[42]. This missense SNP has a CADD score of 31 and is predicted by SIFT to be deleterious and by PolyPhen to be possibly damaging.

Tissue, cell type and pathway enrichment analysis

Using gene expression data for 53 tissues obtained from GTEx [43], we found all brain regions to be significantly enriched for our associated genes with the strongest enrichments for the frontal cortex, BA9 ($P = 2.26 \times 10^{-11}$), the cortex ($P = 8.48 \times 10^{-11}$) and the cerebellar hemisphere ($P = 1.18 \times 10^{-10}$; Figure 4a and Supplementary Table 12). There was no significant enrichment in other tissues of the body. Expression analysis at the cellular level was performed using data sets from the Human Prefrontal cortex by age [44], the Human Cortex [45] and Linnarsson Mouse Brain Atlas [46]. We analysed significant cell types

across datasets, independent cell type associations based on within-dataset conditional analyses and pair-wise cross-datasets conditional analyses (Figure 4b and Supplementary Tables 13 and 14). These analyses identified four neuronal cell types to be enriched for our associated genes. For human data, these were neurons in the cortex (P = 2.16 x 10^{-6}), and GW26 GABAergic neurons in the prefrontal cortex (P = 6.59 x 10^{-8}). For mouse data, these were excitatory glutamatergic neurons in cortical pyramidal layer 5 of the cerebral cortex (TEGLU10; P = 6.98 x 10^{-6}) and excitatory glutamatergic/nitric oxide neurons in the tegmental reticular nucleus of the pons in the hindbrain (HBGLU8; P = 6.74 x 10^{-7}). The enrichment in GABAergic neurons is interesting because there is growing evidence to suggest that impairment of the GABAergic system caused by aging results in an imbalance in the inhibitory/excitatory process involved in the neuronal response to cellular challenges and environmental changes. This results in increased vulnerability to synaptopathy and cognitive decline [47] .

Gene-set analysis performed on curated gene sets and Gene Ontology (GO) [48] terms using the full distribution of SNP P-values from the *Resilience* GWAS identified two GO terms to be significantly enriched after adjustment for multiple testing These were the biological processes“ neuron differentiation” (P = 9.7 x 10^{-07}) and the cellular component “synaptic part” (P = 2.14 x 10^{-06}). Bi-directional conditional analysis using MAGMA 1.08 [49] showed that these two annotations were independent of each other (Supplementary Table 15).

Genetic correlations with other traits

We compared our *Resilience* GWAS with recent published GWAS of cognitive phenotypes, psychiatric and neurological disorders, and global brain imaging phenotypes using LDSR analysis. A moderate negative correlation of *Resilience* with intelligence [42] (r_g= -0.26, P = 1.29 x 10^{-17} and educational attainment [15] (r_g= -0.45, P = 1.64 x 10^{-56} is as expected given that the resilience phenotype was derived from individuals within the UK biobank that has lower than average education years. Of the 13 independent genome-wide significant SNPs for *Resilience*, 6 are associated with intelligence at genome-wide significant levels (P < 5 x 10^{-8}) but the remaining 7 SNPs are not associated with intelligence (P > 0.01). This indicates that some of genetic basis of *Resilience* does not overlap fully with the genetics of intelligence.

When genetic correlation analyses between *Resilience* and psychiatric phenotypes were corrected for multiple testing (P_{bon} < 2.4 x 10^{-3}), *Resilience* had a small positive correlation with unipolar depression [50] (r_g = 0.17, P = 5.0 x 10^{-10}), a small negative correlation with

schizophrenia [51] (r_g = -0.18, P = 1.24 x 10^{-12}) and bipolar disorder [52] (r_g = -0.17, P = 1.84 x 10^{-7}), and a nominally significant negative correlation with neuroticism [53] (r_g= -0.07, P = 2.02 x 10^{-2}). Examination of neurological disorders showed *Resilience* had a small nominally significant correlation with amyotrophic lateral sclerosis (ALS) [54] (r_g = -0.21, P= 1.44 x 10^{-2}), stroke [54] (r_g= 0.08, P = 1.89 x 10^{-2}), and Parkinson's disease [55] (r_g = -0.08, P = 4.58 x 10^{-02}), but Alzheimer's disease (AD) [56] was not significant (r_g = 0.04, P=0.358) (Supplementary Table 16 and Figure 5).

The GWAS of 11 brain phenotypes from the UK Biobank [57] were examined by LDSC for genetic correlation with the *Resilience* (Supplementary Table 17 and Figure 5). The volume of global white and grey matter and cerebral white matter in the left and right hemisphere were examined based on the relationship between brain volumes and cognition [58]. Volume of cerebrospinal fluid was included based of its documented association with brain atrophy [59] and the hippocampus, amygdala and nucleus accumbens were examined as moderators of cognitive function [60] [61]. After adjusting for multiple testing (P_{bon} = 2.4 x 10^{-3}), the only significant correlations found were for white matter volumes where a small positive correlation was found between *Resilience* and global white matter volume (r_g = 0.14, P = 1.19 x 10^{-3}), and the volume of cerebral white matter in the left (r_g = 0.148, P = 1.74 x 10^{-03}) and right hemisphere (r_g = 0.160, P = 7.34 x 10^{-04}).

The correlations of cognitive and psychiatric and neurological disorders are largely supported by gene enrichment analysis of the genes associated with *Resilience* here and previous GWAS of cognitive and psychiatric phenotypes. An analysis of published research from the GWAS catalog [62] showed that the significant SNPs found in this study were previously cited 294 times. A total of 47% of these citations were from studies of cognitive phenotypes (educational attainment, cognitive ability, maths ability and RT) and 5% were from studies of psychiatric disorders (Supplementary Table 18). In addition, when this exercise was repeated for overlapping mapped genes, we found that there was considerable overlap with these phenotypes amongst others. The most significant overlap was where 40 mapped genes in the *Resilience* analysis overlapped with the 99 reported genes for short sleep duration (P= 2.03 x 10^{-57}). In a recent Mendelian randomisation study on sleep duration it was suggested that sleep duration may represent a potential causal pathway for differences in cognitive ability [63] and increase sleep in adults over 60 is associated with poorer cognitive function [64]. There was also a significant overlap with genes associated with extremely high intelligence [65] where 32 *Resilience* mapped genes overlapped with the 81 associated genes reported in

that study (P = 1.17 x 10^{-45}). Many of the overlapping genes for sleep duration and extremely high intelligence were on chromosome 3 (Supplementary Table 19).

Mendelian Randomisation

To investigate whether genetic correlations reflected directional effects, we examined the potential credible causality of the relationship between *Resilience* and phenotypes where independent samples were available using Generalised Summary statistics-based Mendelian Randomisation [66] (GSMR) (Supplementary Table 20 and online methods). We observed a significant bidirectional causal effect of *Resilience* on schizophrenia (bxy = -0.25, P = 7.02 x 10^{-9}) and schizophrenia on *Resilience* (bxy = -0.07, P = 3.80 x 10^{-7}) indicating an inter-relationship between the two phenotypes. By contrast, bipolar disorder and ALS did not have significant credible causality relationships with *Resilience*.

GSMR analysis was also performed using white matter volume variables and *Resilience.* To maintain independence between GWAS datasets, we used the *discovery.Resilience* GWAS that did not include UKB participants with imaging data. The low level of independent significant SNPs in the discovery GWAS did not allow for analysis of the causal effect of *Resilience* on white matter. A nominally significant causal association of white matter volume with *Resilience* was detected (b_{xy} = 0.13, P = 0.049) along with causal associations of left and right cerebral hemisphere white matter volume with *Resilience*. The association with the right hemisphere survived multiple test correction (left: b_{xy} = 0.15, P = 0.005; right: b_{xy} = 0.17, P = 0.002). There is no evidence of substantial pleiotropy in the GSMR analysis.

Examination of the relationship of *Resilience* with RT

Given the strong positive correlation of *Resilience* with RT (r_g = 0.80), a possible concern was that we were just identifying genetic associations with RT that are independent of EY. To examine this further we performed a functional analysis on a GWAS of a dichotomised RT phenotype using all suitable participants in the UK Biobank (n=333,664). This GWAS was perfectly correlated (r_g = 1, P = 7.24 x 10^{-115}) with a previously published GWAS where RT was studied as a quantitative phenotype [29]. We found that while nine of the 13 loci identified in the *Resilience* GWAS overlapped with RT, four loci did not. There was a total of 534 mapped genes for RT and 366 for *Resilience.* Of these, 301 were unique to RT and 133 unique to *Resilience* with 223 shared genes. Only 11 of our 33 prioritized genes in *Resilience*

were among the 27 prioritized genes for RT (Supplementary Figure 5). Pathway enrichment analysis identified GO terms that were enriched for both RT and *Resilience* associated genes (e.g., the cellular component “synaptic part”)) but also showed pathways related to neuronal processes that are only significant in *Resilience* (e.g., the biological process “neuron_differentiation”; Supplementary Table 21). The genetic correlation between *Resilience* and RT is strong because this is an RT-based resilience phenotype. However, there are differences in the associated genes being detected and prioritized because we detected SNPs associated with faster than average RT in individuals that previously showed below average EY, i.e., the resilience phenotype in this study. This phenotype enabled the identification of genetic differences between those individuals in the UK Biobank who preserved or maintained their capability to process information and respond over a 40-year time period compared to individuals who showed diminishing processing speed.

Discussion

This is the first study, to our knowledge, to explore the genetic basis of cognitive resilience in a large data set, here using processing speed measured in later adulthood as a basis for a resilience phenotype. In the absence of longitudinal data, we used a proxy phenotype of EY to measure cognitive performance in earlier adulthood and have combined case-control GWAS with structural equation modelling to extract genetic variants associated with *Resilience* in the UKB. We have shown the robustness of this method by confirming associations detected in a discovery sample replicate in an independent sample. We have successfully identified 13 independent genome-wide significant loci resulting in 366 mapped genes and 33 prioritized genes for *Resilience*. Functional analysis showed significant expression of associated genes in all brain tissues, and particularly in the frontal cortex. Significant enrichment of associated genes was also found at the cellular level in both GABAergic and glutamatergic neurons indicating an excitatory/inhibitory control in the prefrontal cortex, and within biological processes related to neuron differentiation and synaptic activity.

Mapping of GWAS results identified genes that have been previously associated with cognitive decline including *STAU1, SEMF3A, IP6K1, MST1,* the *ATNX2/BRAP* locus, *ALDH2 and DDX27*, where a likely functional missense variant is highly associated. Other associated genes involved with synaptic activity and neurogenesis include *BNS, DAG1, IP6K1* and *TET2,* pointing to potential targets for improvement of cognitive resilience.

One limitation of our study was our reliance on RT to create the *Resilience* phenotype, which results in a strong genetic correlation between *Resilience* with RT. This reflected our study design that detected SNPs associated with faster than average RT or processing speed in individuals that previously showed below average EY. However, the majority of genes prioritized by our *Resilience* GWAS are not prioritized by the RT GWAS and vice versa. We conclude that these findings point to genes that enhance maintenance of processing speed over the life span. Decline in processing speed is a s strong predictor of decline in cognitive processing in older adults [11] and had been found to be associated with cerebral small vessel disease and factors involved in the maintenance of cerebellar morphology[67]. In addition, better cognitive processing speed is also associated with larger cerebral cortex volumes (supporting our finding of a causal relationship with white matter volume), lower levels of inflammatory markers and insulin and is mediated by physical exercise [68]. Over half of our

genome-wide significant loci for *Resilience* are not associated with intelligence, indicating that factors such as reserve, compensation and maintenance may play a role over and above overall intelligence in determining resilience.

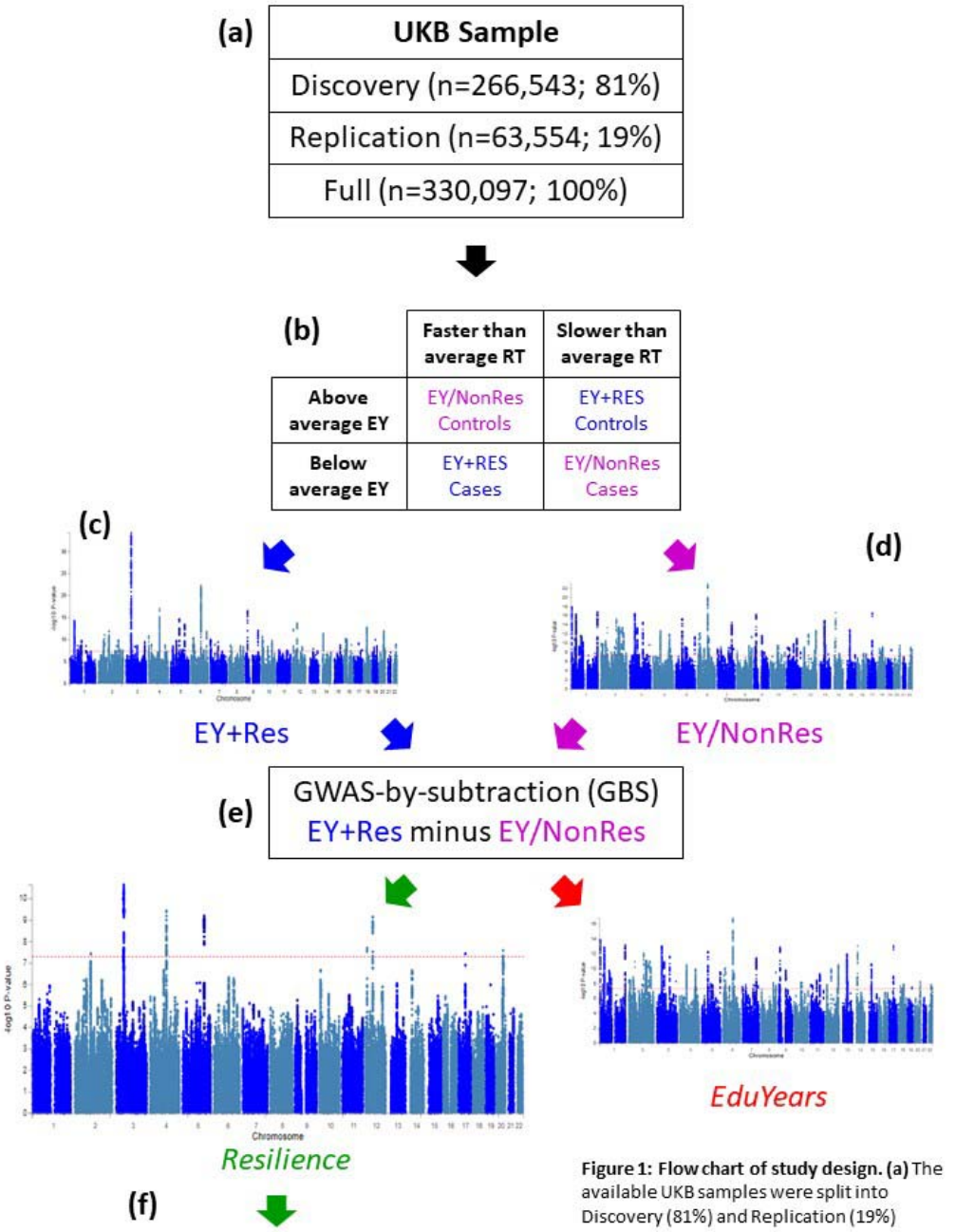
Our use of a proxy phenotype for past cognitive performance and GBS to generate a *Resilience* GWAS was in response to the limitation of not having direct repeated measurements of same cognitive phenotype in large numbers of genotyped samples over an extended time period. Instead, we refined our analysis to make best use of available phenotypic data in UKB to exploit the large sample size available for genetic discovery. We used the proxy phenotype of academic achievement (EY) to represent past cognitive performance in the absence of a direct measure of processing speed. In support of this approach, a study using a sample of 1,560 pupils found that information processing speed is the key predictor of number sense, fluid intelligence and working memory, which in turn predict individual difference in academic achievement [69]. A further limitation is that processing speed as measured by RT is only one component of cognition and it may not be possible to extrapolate the results of this analysis to global cognitive resilience. To support our approach, we explored the use of a much smaller dataset (HRS) which had longitudinal data over time. We showed a genetic correlation between our *Resilience* findings in UKB and cognitive change in HRS. This dataset was not sufficiently powered to confirm the *Resilience* genetic variants. However, this can be addressed by various biobanks that plan new data collection in the future.

This study demonstrated a new method to explore cognitive resilience and identifies associated loci and genes that provide neurobiological insights for this processing speed-based resilience phenotype. It suggests that cognitive resilience is not just a function of superior intelligence and is causally related to variation in white matter volume. This is turn may represent a potential target for studies seeking to enhance resilience therapeutically.

Acknowledgements

This research has been conducted using the UK Biobank Resource (project # 23739). We would like to thank all the participants and scientific researchers who contributed to the data used in this analysis. In additional we would like to acknowledge the support for data analysis provided by Declan Bennett, School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway.

Figures



(f)

Functional analysis

- Description of genetic loci
- Fine mapping of loci
- Gene mapping
- Tissue, cell type and pathway enrichment analysis
- Genetic correlations with other traits
- Mendelian randomisation

Figure 1: Flow chart of study design. (a) The available UKB samples were split into Discovery (81%) and Replication (19%) samples. Following successful replication analysis, the Full sample was also put through the analysis pipeline. **(b)** For Discovery, Replication or Full, samples were assigned to one of four categories based on their EY and RT measures. **(c)** EY+Res cases and controls were analysed in a GWAS. **(d)** EY/NonRes cases and controls were analysed in a GWAS. **(e)** GBS used to subtract the genetic signals for EY/NonRes from EY+Res to result in a *Resilience* GWAS and an *EduYears* GWAS. **(f)** *Resilience* GWAS functionally analysed to identify associated SNPs and genes, and enriched tissues, cell types and pathways, identify genetic correlations with other traits and explore causal relationships between resilience and other traits using Mendelian randomisation

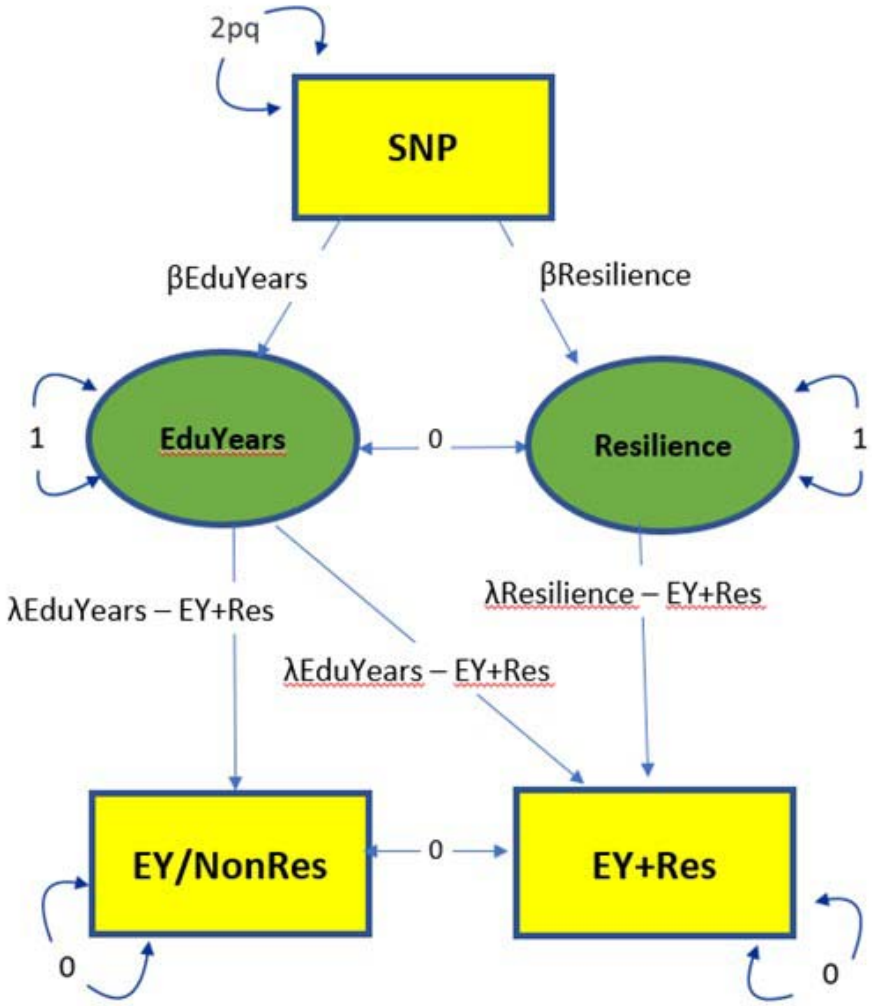
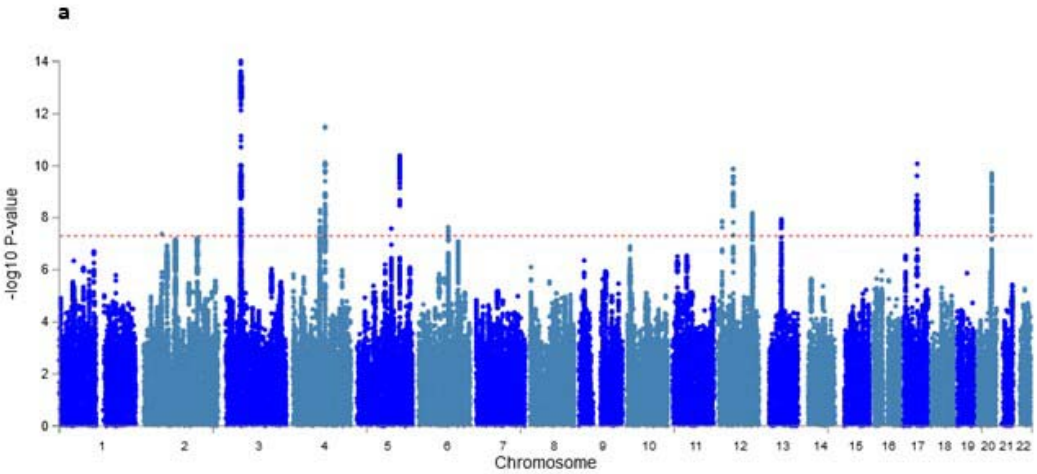
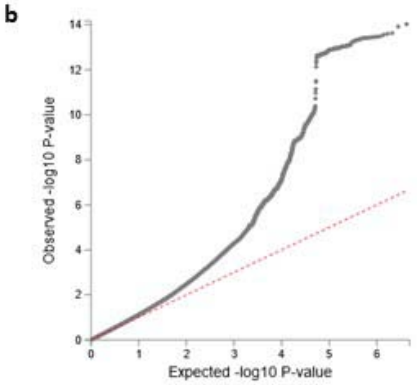
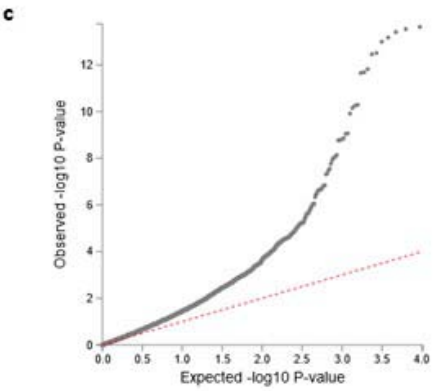
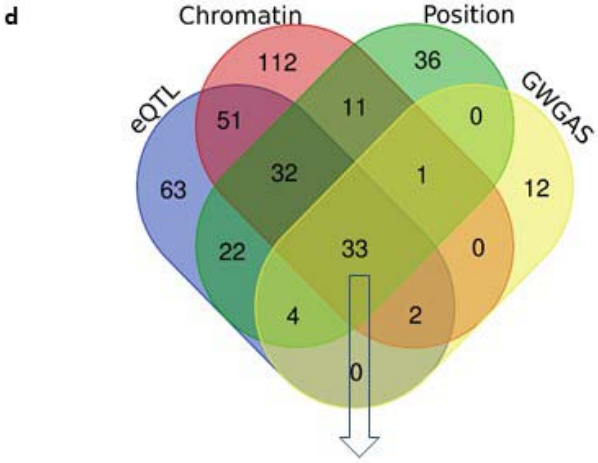


Figure 2: SEM of GWAS-by subtraction model. The observed variables are the GWAS EY+Res and EY/NonRes and SNP and the latent variables (unknown) are *Resilience* and *EduYears*. There are two pathways for the SNPs analysis in this model to EY+Res – the first is through *EduYears* to EY+Res and EY/NonRes and incorporates the genetic effects of the variables used in the phenotype. The other path is through *Resilience* to EY+Res and measures the genetic effect of resilience independent of *EduYears*. To calculate the model, the genetic covariances between EY+Res and EY/NonRes and *Resilience* and *EduYears* are set to 0 and the variances of EY+Res and EY/NonRes are also set to 0. The variance is therefore explained by the latent factors. The SNP value is calculated as 2pq from allele frequencies of the 1000 Genome phase 3 data where p is the reference allele and q the alternative allele.





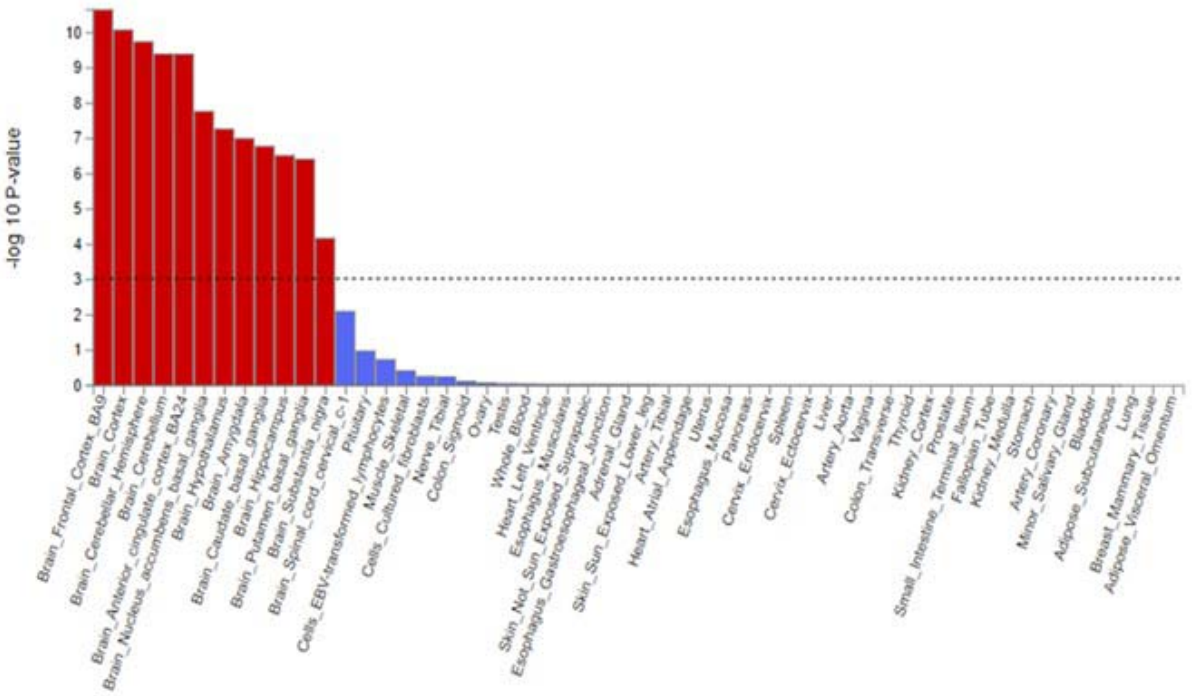




ACAD10, ALDH2, AMT, ATXN2, BRAP, BSN, CAMKV, CSE1L, CTD-2330K9.3,
MAPKAPK5, MON1A,CYSTM1, DAG1, DDN, GNAT1, GPX, IP6K1, MST1
RBM5, RHOA, RNF123, MST1R, NICN1, PFDN1, PRKAG1,
SEMA3F, SH2B3, STAU1, TCTA, TET2, TMEM116, TRAIP, UBA7,

Figure 3: *Resilience* GWAS and gene identification. a Manhattan plot of *Resilience* identifying 13 independent genome-wide significant loci. **b** Quantile – quantile plot of GWAS SNPs. **c** Quantile – quantile plot of the gene-based association test. **d** Venn diagram of overlapping mapped genes by four strategies showing 33 genes were mapped by all four strategies. These genes are listed underneath

(a)



(b)

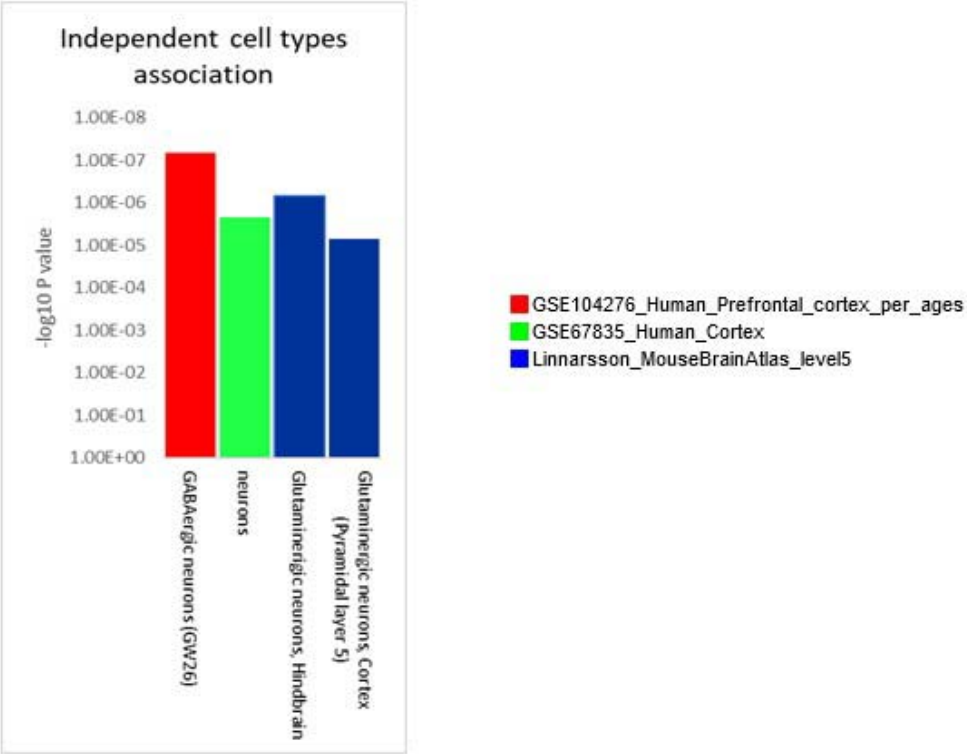


Figure 4: (a) Gene - tissue expression analysis based on GTEx RNA-seq data (b) Independent cell type associations based on within-dataset conditional analyses

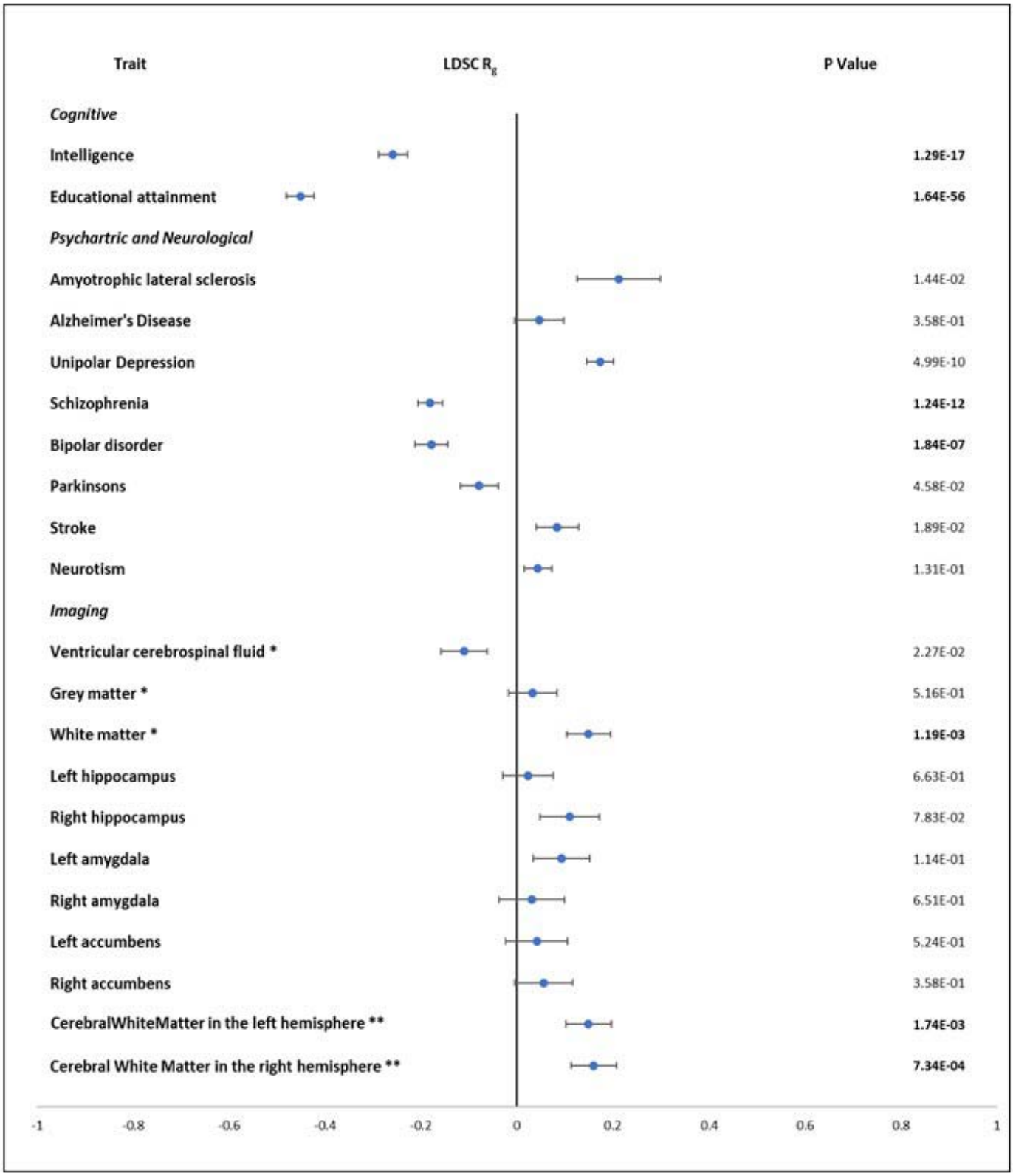


Figure 5: Genetic correlation with cognitive traits , psychiatric and brain disorders and brain imaging, significant P values corrected for multiple testing are in bold.
*normalised for head size , **generated by subcortical volumetric segmentation .

Materials and Methods: (Refer to https://github.com/joanfitz5/cog.res for detailed analysis steps)

Ethics Statement

Our use of the UK Biobank data in this study falls within UK Biobank's generic Research Tissue Bank (RTB) approval from the NHS North West Research Ethics Committee, UK (reference 11/NW/0382).

UK biobank

The UK Biobank (UKB) is a data set of over half a million participants between the ages of 40 and 69, recruited from all over the UK in the period of 2006 to 2010 and had been described extensively elsewhere [15]. We obtained permission to access both the phenotypic and genetic data under project # 23739.

Genetic data: Genotypic data was collected, processed, quality controlled and imputed by UKB [70]. During our in-house quality control of the imputed data, we excluded samples with a Mahalanobis distance >6 SD from multi-mean of European Population structural analysis, removed samples with discordant sex information, chromosomal aneuploidies, high missingness/heterozygosity, retracted consent, and we excluded all related subjects using UKB-provided files on genomic relatedness. The final sample size used in this analysis was 333,664 participants.

Variants were screened by applying quality control filters (geno 0.02, MAF 0.001, info score 0.09 and HWE 0.0001) and removing duplicates resulted in 8,378,152 variants for use in our final analysis.

Phenotypic data

Participants undertook a wide range of cognitive tests. The types of tests and the method of collection and reliability are described elsewhere [71, 72]. Analysis of cross-sectional cognitive data at time zero using IBM SPSS V24 [73] shows a moderate correlation between age and decline in performance on reaction time and a small correlation with numeric memory, pairs matching, prospective memory, and a weak correlation with fluid intelligence (Supplementary Table 22). Fluid intelligence was repeated at two subsequent intervals, however no significant sensitivity to aging was found. Deficiencies in the robustness of the

longitudinal data collected at the second and third time points have been discussed elsewhere [9,10].

Generation of resilience phenotype

Given the lack of longitudinal data, an alternative approach was to use proxy phenotypes. For past cognitive performance we examined the use of educational attainment / years in education [13]. Educational attainment is available for 332,089 individuals in UKB that met our genotypic QC requirements. In the data set, age completed full time education was recorded for participants who did not go to college but not for those who attended higher education. We therefore assigned a default score of 20 to those who attended college and created a binary phenotype using less than or equal to age 17 to divide participants into two categories – above average and below average education years (EY). We then examined the cognitive data and selected the parameter of processing speed as measured by reaction time (RT) as an indicator of current cognitive performance. RT was chosen as it had a good correlation with age and data was available on most participants (N=331,495). RT was adjusted for age and to improve normality [17], the natural log of corrected RT was computed, and a binary RT variable was created using the mean value (5.71). Those with a value less than or equal to the mean were considered to have faster than average processing speed/RT (quicker to react) and those above the mean were considered to have slower than average processing speed/RT. At total of 330,098 individuals had measurements for EY and RT and genetic data and these made up the final sample (Supplementary Table 23).

Using these two binary variables – above or below average EY and faster or slower RT – we created four group of participants (Figure 1b). One of these groups demonstrated high resilience and these were our cases for our first “EY+Res” GWAS who had below average EY previously and faster than average RT now. A second group demonstrated low resilience or cognitive decline, and these were our controls for that GWAS who had above average EY previously and slower than average RT now. The two remaining groups of UKB samples displayed consistent cognitive performance over time. Here our cases for our second “EY/NonRes” GWAS had below average EY previously and slower than average RT now (below average cognition over time) and our controls had above average EY previously and faster than average RT now (above average cognition over time).

GWAS-by subtraction

To extract those SNPs that were associated with resilience only we used Genomics Structural Equation Modelling (GenomicSEM) [22]. There are several processing steps that need to be performed to enable the summary statistics to be processed through GenomicSEM and these are described in the original paper by Grotzinger et al and accompanying tutorials [22, 74]. Following closely the process use by Demange et al [23], we defined a Cholesky model (Figure 2) as follows using the summary statistics from the EY+Res and EY/NonRes GWASs. Both EY+Res and EY/NonRes were regressed on a latent factor, which captured the shared genetic variance in EY (hereafter "*EduYears*"). EY+Res was further regressed on a second latent factor capturing the variance in EY+Res independent of EY/NonRes, hereafter "*Resilience*". Genetic variance in *Resilience* was independent of genetic variance in *EduYears* (r_g = 0) as the *Resilience* factor represents residual genetic variation in our EY+Res phenotype that is not accounted for by the *EduYears* factor. These two latent variables, *Resilience* and *EduYears* were then regressed on each SNP in the original GWASs (EY+Res and EY/NonRes) resulting in new GWAS summary statistics for both *Resilience* and *EduYears* (Figure 2). To calculate the path loadings for λEduYears – EY+Res and λResilience – EY+Res, the model was run without the SNPs.

Execution of GBS

To show replication of our GBS-based *Resilience* GWAS, we divided the UKB into a discovery (81%) and replication (19%) sample. The replication sample included those participants in UKB that had brain imaging data available (n=37,439) and other random participants to give a total sample size of 63,554. The discovery sample consisted of the remaining suitable participants (n= 266,543) Sample sizes used for analysis are shown in Supplementary Table 1. For the discovery, replication and full analysis we performed two initial GWAS for each sample (EY+Res and EY/NonRes) in plink2.0 [75] using sex, age, assessment centre, genotype array and the first 8 principle components of the population stratification analysis as supplied by the UK Biobank. We then performed GBS on both sets of samples resulting in a *discovery.Resilience* and *discovery.EduYears* GWAS, a *replication.Resilience* and a *replication.EduYears* GWAS, and later a full *Resilience* GWAS and *EduYears* GWAS.

Calculation of sample size after GBS

Running the analysis through GBS alters the sample size and it is necessary to calculate the new value for downstream analysis. To calculate sample size or effective N (Neff) of the *Resilience* GWAS for test, replication and full we followed the procedure specified in GenomicSEM[22, 76] and by Demange et al [17] (see URLs). To do this we needed to determine path loading for the models used in the three analysis as the path loading differs with different sample sizes. We trimmed our data to only include SNPs with a MAF of >0.10 and <0.40 as low and high MAF can bias the result. The output of this analysis and the calculations of sample size is in Supplementary Table 24.

Additional GWAS

To examine the effect of GBS on EY+Res we ran two further case-control GWAS of above/below average EY and faster/slower than average age corrected RT. We mirrored the sample size used to generate EY+Res and EY/NonRes by randomly selecting 82,000 samples as cases and controls from the data set. This analysis was run in plink 2.0 [75] using sex, age, assessment centre, genotype array and the first 8 principle components of the population stratification analysis as supplied by the UKB. To further interpret the relationship of *resilience* to RT we performed a quantitative GWAS on the 164,000 samples used in the cases control study. In addition, we also performed a full dichotomised study of all 333,664 individuals in the UK biobank with RT data.

Identification of genomic loci associated with *Resilience*

Manhattan plots of GWAS outputs from original phenotypes and GBS outputs were generated in FUMA v 1.3.6 [24] using a P-value setting of $<5 \times 10^{-8}$ for genome-wide significant SNPs. We used an LD r^2 setting of 0.6 and the 1000G phase 3 European reference panel to identify independent lead SNPs and an additional r^2 setting of 0.1 to identify lead SNPs and a maximum distance for LD blocks of 250 kb to separate findings into separate genetic loci. Conditional analysis was performed where there was more than one independent significant SNP within 1000 kb distance using --condition command in Plink 1.9 [75], which adds a SNP as a covariate in GWAS analysis.

FINEMAP [28] was used to investigate causal SNPs by analysing the relationship between the candidate GWAS SNPs generated in FUMA and LD data. LD files were generated in plink 1.9 using the --r square spaces command. Results of SNPs listed by Bayes Factor for each locus were examined as well as the configuration files generated by FINEMAP to examine for causal SNPs sets. The maximum number of SNPs in a set was fixed at 3.

Function analysis of GWAS output

We used FUMA v 1.3.6 [24] to perform functional analysis. We used the default settings as described in the Tutorial section of the website and in previous publications [56, 77] . FUMA analysis of *Resilience* is published and can be viewed publicly in FUMA as ID:171. We used the calculated effective sample size of 111,316 (Neff) for the analysis of the *Resilience* output to examine the functional consequences of SNPs on genes, Combined Annotation Dependent Depletion (CADD) scores, chromatin states and Regulome DB analysis.

Mapping SNPs to genes

Gene-mapping was performed in FUMA using three strategies: (a) *Positional mapping* which mapped SNPs to genes based on their genomic location within a 10 kb window of known gene boundaries. (b) *Expression quantitative trait (eQTL) mapping* which aligned cis-eQTL SNPs to genes whose expression they affected, selecting information from tissue types in 4 data sets in FUMA (PsychENCODE [78], BIOS QTL [79], Blood eQTL [80], and GTEx 8 [81]). (c) *Chromatin interaction mapping* using the 3D DNA to DNA interactions mapped SNPs to genes.

Gene-set analyses: The GENE2FUNC function within FUMA examines enrichment of mapped genes using hypergeometic tests of 9494 gene-sets form GTEx [82] , MSigDB [83] and GWAS catalog [84].

MAGMA gene-based analysis

FUMA computes a gene-based genome-wide association analysis (GWGAS) from the SNP-based P-value from the GWAS. A total of 18,879 protein coding genes containing a minimum of one GWAS SNP were used in this analysis and were used to test for association with 53 tissue types. Associations were Bonferroni corrected for multiple testing with $P < 0.05/18{,}879 = 2.648 \times 10^{-6}$.

We further explored the sets of associated genes in cell type specificity analyses with scRNA-seq in FUMA [85] using the following data sets: GSE104276 Human Prefrontal cortex per ages [44], GSE67835 Human Cortex [86] and Linnarsson Mouse Brain Atlas [46]. We analysed significant cell types across datasets, independent cell type associations based on within-dataset conditional analyses and pair-wise cross-datasets conditional analyses.

Comparison with published traits

LD score regression (LDSR) analysis was performed using the LDSC function within GenomicSEM [22] to examine the genetic correlation between *Resilience* with other phenotypes. Various sources were used to obtain summary statistics from GWAS of published research in psychiatry, brain imaging, and other traits of interest (supplementary Table 16 and 17). Munged summary statistic files generated during GBS were used for *Resilience*, *EduYears*, *EY+Res* and *EY/NonRes* in the LDSR. Associations were Bonferroni corrected for multiple testing with $P < 0.05/21 = 2.88 \times 10^{-3}$.

Mendelian Randomisation

Mendelian randomisation was performed using Generalized Summary statistics-based Mendelian Randomization [87] GSMR using the GCTA tool [88]. The procedure examines credible causal associations between different traits based on GWAS outputs and requires non-overlapping samples. This restricted our analysis because most of the traits examined by LDSC contained UKB participants. However the sample used for the *discovery.Resilience* GWAS (section 1.2.1) does not contain individuals that have imaging data within the UKB so we used this cohort to examine unidirectional and bidirectional causal associations between *Resilience* and phenotypes that showed significant correlations with *Resilience* using LDSC. We used a HEIDI-outlier p-value of 0.01 for outlier detection analysis. Given the low level of independent significant SNPs in the *discovery.Resilience* GWAS and the imaging GWAS, we reduced the default minimum level of significant SNPs from 10 to 8. For the disorders of ALS, bipolar disorder and schizophrenia we used the full *Resilience* GWAS and ran the analysis at the default setting of a minimum of 10. Associations were Bonferroni corrected for multiple testing with $P < 0.05/12 = 4.23 \times 10^{-3}$.

US Health and Retirement Study (HRS)

The HRS is a longitudinal study of adults aged 50 years or older in households in the United States. The study commenced in 1992 and participants were interviewed at baseline and every two subsequent years. Phenotypic data was accessed through the HRS web site (see URLs). Genetic data was supplied through dbGaP as project 18937.

Genetic data

The imputed genetic data were available per chromosome as probability files (gprob.gz). Pgen files were created and were screened by applying quality control filters (geno 0.02, MAF 0.01, info score 0.09 and HWE 0.000001) and removing duplicates for each chromosome. In addition, SNPs in the kgp format were converted to RSID using the USCS browser [89] .

Cognitive phenotypes

Several cognitive tests were administered by interviewers, by phone or face-to-face. These tests are described in detail on the HRS website. The HRS dataset does not contain a reaction time or processing speed measure. To maximise the number of participants with two data points, we consisted of the interviewer reading a randomised list of 10 nouns to the respondent from one of four lists, and afterwards asked the respondent to recall as many words as possible.

Data was available for eleven intervals, each two years apart, representing a time span of 22 years. Cognitive data was available for 38,183 participants. Of these, 15,620 had genetic data and this was reduced further when individuals were eliminated based on ethnicity (non-Caucasian), missing data or less than 2 data points, to reduce the available sample to 9,526 individuals. Linear mixed modelling (LMM) in SPSS was used to analysis the data as this method allows for missing data in longitudinal samples, using the covariates of gender, birthyear, education, time of test and quadratic time point [26] .

GWAS of cognitive change over time:

We created a cognitive phenotype for cognitive change from the LMM output and performed a GWAS using the first six PCA components as covariates as recommended in the genotype QC report [25]. The output of the GWAS was processes through FUMA.

URLs:

UK Biobank: http://biobank.ndph.ox.ac.uk

Plink: www.cog-genomics.org/plink/2.0/

GenomicSEM: https://github.com/MichelNivard/GenomicSEM/wiki

GWAS-by-subtraction: https://rpubs.com/MichelNivard/565885

GBS sample size (N effective) calculation: https://github.com/PerlineDemange/non-cognitive/blob/master/GenomicSEM/Cholesky%20model/Calculation_samplesize.R

Functional Mapping and Annotation (FUMA): https://fuma.ctglab.nl/downpage.html

Venn diagram: http://bioinformatics.psb.ugent.be/webtools/Venn/

FINEMAP: http://www.christianbenner.com/

Generalised Summary-data-based MR: https://cnsgenomics.com/software/gcta/#GSMR

GWAS Catalog: https://www.ebi.ac.uk/gwas/

Brain phenotypes: https://open.win.ox.ac.uk/ukbiobank/big40/

GWAS Atlas: https://atlas.ctglab.nl/

Ensembl Variant Effect Predictor: https://www.ensembl.org/Tools/VEP

HRS: https://hrsonline.isr.umich.edu/

references

1. Deary IJ, Corley J, Gow AJ, Harris SE, Houlihan LM, Marioni RE et al. Age-associated cognitive decline. British Medical Bulletin. 2009;92(1):135-52. doi:10.1093/bmb/ldp033.

2. Andrews SJ, Das D, Cherbuin N, Anstey KJ, Easteal S. Association of genetic risk factors with cognitive decline: the PATH through life project. Neurobiol Aging. 2016;41:150-8. doi:10.1016/j.neurobiolaging.2016.02.016.

3. World Health O. World report on ageing and health. Geneva: World Health Organization; 2015.

4. Staal MA, Bolton A, Yaroush R, Bourne Jr L. Cognitive performance and resilience to stress. Biobehavioral resilience to stress. 2008:259-99.

5. Cabeza R, Albert M, Belleville S, Craik FIM, Duarte A, Grady CL et al. Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. Nat Rev Neurosci. 2018;19(11):701-10. doi:10.1038/s41583-018-0068-2.

6. Stern Y, Arenaza-Urquijo EM, Bartrés-Faz D, Belleville S, Cantilon M, Chetelat G et al. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. Alzheimer's & Dementia. 2018. doi:https://doi.org/10.1016/j.jalz.2018.07.219.

7. Cadar D, Robitaille A, Pattie A, Deary I, Muniz-Terrera G. The Long Arm of Childhood Intelligence on Terminal Decline: Evidence From the Lothian Birth Cohort 1921. Psychol Aging. 2020;35(6):806-17. doi:10.1037/pag0000477.

8. Guerra-Carrillo B, Katovich K, Bunge SA. Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. PLOS ONE. 2017;12(8):e0182276. doi:10.1371/journal.pone.0182276.

9. Lövdén M, Fratiglioni L, Glymour MM, Lindenberger U, Tucker-Drob EM. Education and Cognitive Functioning Across the Life Span. Psychological Science in the Public Interest. 2020;21(1):6-41. doi:10.1177/1529100620920576.

10. Geary DC. The Spark of Life and the Unification of Intelligence, Health, and Aging. Current Directions in Psychological Science. 2019;28(3):223-8. doi:10.1177/0963721419829719.

11. Salthouse T. The processing-speed theory of adult age differences in cognition. Psychological review. 1996;103 3:403-28.

12. Salthouse TA, Ferrer-Caja E. What needs to be explained to account for age-related effects on multiple cognitive variables? Psychology and aging. 2003;18(1):91.

13. Schubert A-L, Nunez MD, Hagemann D, Vandekerckhove J. Individual Differences in Cortical Processing Speed Predict Cognitive Abilities: a Model-Based Cognitive Neuroscience Account. Computational Brain & Behavior. 2019;2(2):64-84. doi:10.1007/s42113-018-0021-5.

14. Scheiber C, Chen H, Kaufman AS, Weiss LG. How Much Does WAIS-IV Perceptual Reasoning Decline Across the 20 to 90-Year Lifespan When Processing Speed is Controlled? Applied Neuropsychology: Adult. 2017;24(2):116-31. doi:10.1080/23279095.2015.1107564.

15. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J 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 Med. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779.

16. Deary IJ. Intelligence. Current Biology. 2013;23(16):R673-R6. doi:https://doi.org/10.1016/j.cub.2013.07.021.

17. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Mol Psychiatry. 2016;21(6):758-67. doi:10.1038/mp.2016.45.

18. Plomin R, von Stumm S. The new genetics of intelligence. Nature Reviews Genetics. 2018;19:148. doi:10.1038/nrg.2017.104.

19. Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(38):13790-4. doi:10.1073/pnas.1404623111.

20. Gordon R, Smith-Spark JH, Newton EJ, Henry LA. Executive Function and Academic Achievement in Primary School Children: The Use of Task-Related Processing Speed. Frontiers in Psychology. 2018;9(582). doi:10.3389/fpsyg.2018.00582.

21. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics. 2018;50(8):1112-21. doi:10.1038/s41588-018-0147-3.

22. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, David HW et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nature Human Behaviour. 2019;3(5):513-25. doi:http://dx.doi.org/10.1038/s41562-019-0566-x.

23. Demange PA, Malanchini M, Mallard TT, Biroli P, Cox SR, Grotzinger AD et al. Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. Nature Genetics. 2021;53(1):35-44. doi:10.1038/s41588-020-00754-2.

24. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5.

25. Faul J, Smith J, Zhao W. Health and retirement study: candidate gene and SNP data description. Health Retire Study, Univ Mich, Ann Arbor, MI http://hrsonline isr umich edu/sitedocs/genetics/candidategene/CandidateGeneSNPDataDescription Google Scholar Article Location. 2014.

26. Field A. Discovering statistics using IBM SPSS statistics. sage; 2013.

27. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014;46(3):310-5. doi:10.1038/ng.2892.

28. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493-501. doi:10.1093/bioinformatics/btw018.

29. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nature Communications. 2018;9(1):2098. doi:10.1038/s41467-018-04362-x.

30. Tasaki S, Gaiteri C, Mostafavi S, Yu L, Wang Y, De Jager PL et al. Multi-omic Directed Networks Describe Features of Gene Regulation in Aged Brains and Expand the Set of Genes Driving Cognitive Decline. Front Genet. 2018;9:294-. doi:10.3389/fgene.2018.00294.

31. Annamneedi A, Caliskan G, Müller S, Montag D, Budinger E, Angenstein F et al. Ablation of the presynaptic organizer Bassoon in excitatory neurons retards dentate gyrus maturation and enhances learning performance. Brain Struct Funct. 2018;223(7):3423-45. doi:10.1007/s00429-018-1692-3.

32. Kim M-G, Zhang S, Park H, Park SJ, Kim S, Chung C. Inositol hexakisphosphate kinase-1 is a key mediator of prepulse inhibition and short-term fear memory. Mol Brain. 2020;13(1):72-. doi:10.1186/s13041-020-00615-3.

33. Wang P, Geng J, Gao J, Zhao H, Li J, Shi Y et al. Macrophage achieves self-protection against oxidative stress-induced ageing through the Mst-Nrf2 axis. Nat Commun. 2019;10(1):755-. doi:10.1038/s41467-019-08680-6.

34. Gontier G, Iyer M, Shea JM, Bieri G, Wheatley EG, Ramalho-Santos M et al. Tet2 Rescues Age-Related Regenerative Decline and Enhances Cognitive Function in the Adult Mouse Brain. Cell Rep. 2018;22(8):1974-81. doi:10.1016/j.celrep.2018.02.001.

35. Gardiner SL, Trompet S, Sabayan B, Boogaard MW, Jukema JW, Slagboom PE et al. Repeat variations in polyglutamine disease–associated genes and cognitive function in old age. Neurobiology of Aging. 2019;84:236.e17-.e28. doi:https://doi.org/10.1016/j.neurobiolaging.2019.08.002.

36. Lastres-Becker I, Nonis D, Nowock J, Auburger G. New alternative splicing variants of the ATXN2 transcript. Neurological Research and Practice. 2019;1(1):22. doi:10.1186/s42466-019-0025-1.

37. Timmers PR, Mounier N, Lall K, Fischer K, Ning Z, Feng X et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. Elife. 2019;8:e39856. doi:10.7554/eLife.39856.

38. Melzer D, Pilling LC, Ferrucci L. The genetics of human ageing. Nature Reviews Genetics. 2020;21(2):88-101. doi:10.1038/s41576-019-0183-6.

39. Wu NN, Ren J. Aldehyde Dehydrogenase 2 (ALDH2) and Aging: Is There a Sensible Link? Advances in experimental medicine and biology. 2019;1193:237-53. doi:10.1007/978-981-13-6260-6_15.

40. Jackson TC, Kotermanski SE, Kochanek PM. Whole-transcriptome microarray analysis reveals regulation of Rab4 by RBM5 in neurons. Neuroscience. 2017;361:93-107. doi:10.1016/j.neuroscience.2017.08.014.

41. Panzanelli P, Früh S, Fritschy J-M. Differential role of GABAA receptors and neuroligin 2 for perisomatic GABAergic synapse formation in the hippocampus. Brain Structure and Function. 2017;222(9):4149-61. doi:10.1007/s00429-017-1462-7.

42. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016;17(1):122. doi:10.1186/s13059-016-0974-4.

43. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015;348(6235):648. doi:10.1126/science.1262110.

44. Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature. 2018;555(7697):524-8. doi:10.1038/nature25980.

45. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM et al. A survey of human brain transcriptome diversity at the single cell level. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(23):7285-90. doi:10.1073/pnas.1507125112.

46. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J et al. Molecular Architecture of the Mouse Nervous System. Cell. 2018;174(4):999-1014.e22. doi:10.1016/j.cell.2018.06.021.

47. Rozycka A, Liguz-Lecznar M. The space where aging acts: focus on the GABAergic synapse. Aging Cell. 2017;16(4):634-43. doi:10.1111/acel.12605.

48. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25(1):25-9. doi:10.1038/75556.

49. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLOS Computational Biology. 2015;11(4):e1004219. doi:10.1371/journal.pcbi.1004219.

50. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nature Genetics. 2018;50(7):920-7. doi:10.1038/s41588-018-0151-7.

51. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nature genetics. 2018;50(3):381-9. doi:10.1038/s41588-018-0059-2.

52. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature Genetics. 2019;51(5):793-803. doi:10.1038/s41588-019-0397-8.

53. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shirali M et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nature Genetics. 2018;50(1):6-11. doi:10.1038/s41588-017-0013-8.

54. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nature genetics. 2016;48(9):1043-8. doi:10.1038/ng.3622.

55. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. The Lancet Neurology. 2019;18(12):1091-102. doi:10.1016/S1474-4422(19)30320-5.

56. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nature Genetics. 2019;51(3):404-13. doi:10.1038/s41588-018-0311-9.

57. Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K et al. Enhanced Brain Imaging Genetics in UK Biobank. bioRxiv. 2020:2020.07.27.223545. doi:10.1101/2020.07.27.223545.

58. Nave G, Jung WH, Karlsson Linnér R, Kable JW, Koellinger PD. Are Bigger Brains Smarter? Evidence From a Large-Scale Preregistered Study. Psychol Sci. 2019;30(1):43-54. doi:10.1177/0956797618808470.

59. Orellana C, Ferreira D, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M et al. Measuring Global Brain Atrophy with the Brain Volume/Cerebrospinal Fluid Index: Normative Values, Cut-Offs and Clinical Associations. Neuro-degenerative diseases. 2016;16(1-2):77-86. doi:10.1159/000442443.

60. Lisman J, Buzsáki G, Eichenbaum H, Nadel L, Ranganath C, Redish AD. Viewpoints: how the hippocampus contributes to memory, navigation and cognition. Nat Neurosci. 2017;20(11):1434-47. doi:10.1038/nn.4661.

61. Floresco SB. The nucleus accumbens: an interface between cognition, emotion, and action. Annual review of psychology. 2015;66:25-52. doi:10.1146/annurev-psych-010213-115159.

62. Buniello A, MacArthur JA L, Cerezo M, Harris LW, Hayhurst J, Malangone C et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. 2019;47(D1):D1005-D12. doi:10.1093/nar/gky1120.

63. Henry A, Katsoulis M, Masi S, Fatemifar G, Denaxas S, Acosta D et al. The relationship between sleep duration, cognition and dementia: a Mendelian randomization study. Int J Epidemiol. 2019;48(3):849-60. doi:10.1093/ije/dyz071.

64. Low DV, Wu MN, Spira AP. Sleep Duration and Cognition in a Nationally Representative Sample of U.S. Older Adults. The American Journal of Geriatric Psychiatry. 2019;27(12):1386-96. doi:https://doi.org/10.1016/j.jagp.2019.07.001.

65. Coleman JRI, Bryois J, Gaspar HA, Jansen PR, Savage JE, Skene N et al. Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. Molecular psychiatry. 2019;24(2):182-97. doi:10.1038/s41380-018-0040-6.

66. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun. 2018;9(1):224. doi:10.1038/s41467-017-02317-2.

67. Eckert MA. Slowing Down: Age-Related Neurobiological Predictors of Processing Speed. Frontiers in Neuroscience. 2011;5:25. doi:10.3389/fnins.2011.00025.

68. Bott NT, Bettcher BM, Yokoyama JS, Frazier DT, Wynn M, Karydas A et al. Youthful Processing Speed in Older Adults: Genetic, Biological, and Behavioral Predictors of Cognitive Processing Speed Trajectories in Aging. Frontiers in Aging Neuroscience. 2017;9(55). doi:10.3389/fnagi.2017.00055.

69. Tikhomirova T, Malykh A, Malykh S. Predicting Academic Achievement with Cognitive Abilities: Cross-Sectional Study across School Education. Behavioral Sciences. 2020;10(10):158.

70. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9. doi:10.1038/s41586-018-0579-z.

71. Lyall DM, Cullen B, Allerhand M, Smith DJ, Mackay D, Evans J et al. Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. PLoS One. 2016;11(4):e0154222. doi:10.1371/journal.pone.0154222.

72. Fawns-Ritchie C, Deary IJ. Reliability and validity of the UK Biobank cognitive tests. PloS one. 2020;15(4):e0231627-e. doi:10.1371/journal.pone.0231627.

73. IBM Corp. Released 2016. IBM SPSS Statistics for Windows VA, NY: IBM Corp.

74. Nivard M. GenomicSem. 2019. https://github.com/MichelNivard/GenomicSEM/wiki.

75. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1). doi:10.1186/s13742-015-0047-8.

76. Mallard TT, Linnér RK, Grotzinger AD, Sanchez-Roige S, Seidlitz J, Okbay A et al. Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. bioRxiv. 2020:603134. doi:10.1101/603134.

77. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature Genetics. 2018;50(7):912-9. doi:10.1038/s41588-018-0152-6.

78. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP et al. Comprehensive functional genomic resource and integrative model for the human brain. Science. 2018;362(6420). doi:10.1126/science.aat8464.

79. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nature Genetics. 2017;49(1):131-8. doi:10.1038/ng.3721.

80. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature Genetics. 2013;45(10):1238-43. doi:10.1038/ng.2756.

81. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-13. doi:10.1038/nature24277.

82. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreservation and Biobanking. 2015;13(5):311-9. doi:10.1089/bio.2015.0032.

83. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25. doi:10.1016/j.cels.2015.12.004.

84. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-d12. doi:10.1093/nar/gky1120.

85. Watanabe K, Umićević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. Nat Commun. 2019;10(1):3222. doi:10.1038/s41467-019-11181-1.

86. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM et al. A survey of human brain transcriptome diversity at the single cell level. Proceedings of the National Academy of Sciences. 2015;112(23):7285. doi:10.1073/pnas.1507125112.

87. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun. 2018;9(1):224. doi:10.1038/s41467-017-02317-2.

88. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011.

89. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM et al. The human genome browser at UCSC. Genome research. 2002;12(6):996-1006. doi:10.1101/gr.229102.

Figure Legends

Figure 1: Flow chart of study design. (a) The available UKB samples were split into Discovery (81%) and Replication (19%) samples. Following successful replication analysis, the Full sample was also put through the analysis pipeline. (b) For Discovery, Replication or Full, samples were assigned to one of four categories based on their EY and RT measures. (c) EY+Res cases and controls were analysed in a GWAS. (d) EY/NonRes cases and controls were analysed in a GWAS. (e) GBS used to subtract the genetic signals for EY/NonRes from EY+Res to result in a Resilience GWAS and an EduYears GWAS. (f) Resilience GWAS functionally analysed to identify associated SNPs and genes, and enriched tissues, cell types and pathways, identify genetic correlations with other traits and explore causal relationships between resilience and other traits using Mendelian randomisation.

Figure 2: SEM of GWAS-by subtraction model. The observed variables are the GWAS EY+Res and EY/NonRes and SNP and the latent variables (unknown) are Resilience and EduYears. There are two pathways for the SNPs analysis in this model to EY+Res – the first is through EduYears to EY+Res and EY/NonRes and incorporates the genetic effects of the variables used in the phenotype. The other path is through Resilience to EY+Res and measures the genetic effect of resilience independent of EduYears. To calculate the model, the genetic covariances between EY+Res and EY/NonRes and Resilience and EduYears are set to 0 and the variances of EY+Res and EY/NonRes are also set to 0. The variance is therefore explained by the latent factors. The SNP value is calculated as 2pq from allele frequencies of the 1000 Genome phase 3 data where p is the reference allele and q the alternative allele.

Figure 3: *Resilience* GWAS and gene identification. (a) Manhattan plot of *Resilience* identifying 13 independent genome-wide significant loci. (b) Quantile – quantile plot of GWAS SNPs. (c) Quantile –quantile plot of the gene-based association test. (d) Venn diagram of overlapping mapped genes by four strategies showing 33 genes were mapped by all four strategies. These genes are listed underneath.

Figure 4: Tissue and cell type expression analysis. (a) Gene - tissue expression analysis based on GTEx RNA-seq data. Results that are still significant after correction for multiple testing are in red. (b) Independent cell type associations based on within-dataset conditional analyses.

Figure 5: Genetic correlation with cognitive traits, psychiatric and brain disorders, and brain imaging phenotypes. Significant P values corrected for multiple testing are in bold, *normalised for head size ,**generated by subcortical volumetric segmentation.