

IMMEDIATE COMMUNICATION

Genome-wide association study of cognitive functions and educational attainment in UK Biobank ($N = 112\,151$)

G Davies^{1,2,11}, RE Marioni^{1,3,4,11}, DC Liewald^{1,2}, WD Hill^{1,2}, SP Hagenaars^{1,2,5}, SE Harris^{1,3}, SJ Ritchie^{1,2}, M Luciano^{1,2}, C Fawns-Ritchie^{1,2}, D Lyall⁶, B Cullen⁶, SR Cox^{1,2}, C Hayward⁷, DJ Porteous^{1,3}, J Evans⁶, AM McIntosh^{1,5}, J Gallacher⁸, N Craddock⁹, JP Pell⁶, DJ Smith⁶, CR Gale^{1,2,10} and IJ Deary^{1,2}

People's differences in cognitive functions are partly heritable and are associated with important life outcomes. Previous genome-wide association (GWA) studies of cognitive functions have found evidence for polygenic effects yet, to date, there are few replicated genetic associations. Here we use data from the UK Biobank sample to investigate the genetic contributions to variation in tests of three cognitive functions and in educational attainment. GWA analyses were performed for verbal–numerical reasoning ($N = 36\,035$), memory ($N = 112\,067$), reaction time ($N = 111\,483$) and for the attainment of a college or a university degree ($N = 111\,114$). We report genome-wide significant single-nucleotide polymorphism (SNP)-based associations in 20 genomic regions, and significant gene-based findings in 46 regions. These include findings in the *ATXN2*, *CYP2DG*, *APBA1* and *CADM2* genes. We report replication of these hits in published GWA studies of cognitive function, educational attainment and childhood intelligence. There is also replication, in UK Biobank, of SNP hits reported previously in GWA studies of educational attainment and cognitive function. GCTA-GREML analyses, using common SNPs (minor allele frequency > 0.01), indicated significant SNP-based heritabilities of 31% (s.e.m. = 1.8%) for verbal–numerical reasoning, 5% (s.e.m. = 0.6%) for memory, 11% (s.e.m. = 0.6%) for reaction time and 21% (s.e.m. = 0.6%) for educational attainment. Polygenic score analyses indicate that up to 5% of the variance in cognitive test scores can be predicted in an independent cohort. The genomic regions identified include several novel loci, some of which have been associated with intracranial volume, neurodegeneration, Alzheimer's disease and schizophrenia.

Molecular Psychiatry (2016) **21**, 758–767; doi:10.1038/mp.2016.45; published online 5 April 2016

INTRODUCTION

Cognitive functions have important roles in human mental and physical well-being. Better cognitive function in youth is associated with lower risk of some psychiatric disorders and physical illness later in the life course, and with reduced mortality risk.^{1,2} The reverse is also true; some mental and physical illnesses are associated with a lowering of cognitive capabilities in youth and over the life course.^{3,4} Higher cognitive ability in youth is associated also with higher educational attainment and adult social position.⁵ Domains of cognitive functioning differ in their associations with ageing; some have trajectories of decline (for example, processing speed and some types of memory), whereas others (for example, knowledge-based tests) hold their levels better over the adult life course.^{6,7} Therefore, it is important to understand the causes of people's differences in cognitive functions.

One source of cognitive differences is genetic variation. Cognitive functions have a substantial heritability. This has been found by using twin and family studies,^{8–12} and by molecular genetic methods, such as Genome-wide Complex Trait Analysis

(GCTA-GREML),^{13,14} which estimates heritability based on common single-nucleotide polymorphisms (SNPs).

Some explanation is required regarding cognitive phenotypes. All tests of cognitive ability correlate positively, though not perfectly; that is, people who do well on one type of cognitive test tend to do well on the others.¹⁵ It is this regularity that is the basis for the construct of general cognitive ability, which is usually abbreviated to *g*. There are also separable domains of cognitive functioning.^{10,15} Differences in individual cognitive test score performances may be due to: (1) differences in general cognitive function described by the variance shared by all cognitive domains, *g*; (2) differences in test performance specific to a cognitive domain; and (3) differences specific to a particular test.¹⁶ Twin and SNP-based GCTA-GREML studies have found that there is substantial heritability for general cognitive function, and also some heritability for cognitive domains and specific cognitive skills.^{17,18} They also find that there are significant genetic correlations among tests of different cognitive domains, and between cognitive abilities and education, which also shows substantial heritability.^{18,19}

¹Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; ²Department of Psychology, University of Edinburgh, Edinburgh, UK; ³Medical Genetics Section, University of Edinburgh Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK; ⁴Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia; ⁵Division of Psychiatry, University of Edinburgh, Edinburgh, UK; ⁶Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK; ⁷MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; ⁸Department of Psychiatry, University of Oxford, Oxford, UK; ⁹Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK and ¹⁰MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK. Correspondence: Professor IJ Deary, Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.
E-mail: i.deary@ed.ac.uk

¹¹These authors contributed equally to this work.

Received 24 August 2015; revised 14 January 2016; accepted 11 February 2016; published online 5 April 2016

Genome-wide association studies (GWAS) of cognitive functions have been successful in estimating SNP-based heritability, and in using summary GWAS data to make predictions of cognitive phenotypes in independent samples.¹³ However, they have been less successful in identifying the specific genetic variants that cause cognitive differences. The largest studies to date have been the CHARGE-Cognitive Working Group's studies^{13,20,21} and those on educational phenotypes by the Social Science Genetics Association Consortium.^{22–24} In a study of 53 949 individuals with data on general cognitive function, there were three genome-wide significant hits in three genomic regions, with the closest genes being *APOE/TOMM40*, *AKAP6* and *MIR2113*.¹³ In a study of 32 070 individuals with data on processing speed (mostly digit-symbol substitution-type tests) there was one genome-wide significant hit, near *CADM2*.²¹ In a study of 29 076 individuals with data on verbal declarative memory there were three genome-wide significant hits, near *APOE* and genes associated with immune response.²⁰

Given the large phenotypic and genetic correlations between general cognitive ability and educational attainment,^{18,25,26} others have used the latter as a proxy phenotype for cognitive ability.²³ This has led to finding three independent genome-wide significant variants, with the nearest genes being *LOC100129185*, *LRRN2* and *LOC150577*.^{23,24} The latter two hits were with the phenotype of having attained a college or university degree, which is the educational phenotype used in the present study. Genome-wide SNPs from GWAS analyses of education phenotypes identified variants related to cognitive performance phenotypes, implicating the genes *KNCMA1*, *NRXN1*, *POU2F3* and *SCRT*.²³ There has been replication of a SNP (rs1906252) that influences both education and general cognitive function.^{13,27}

GWAS meta-analytic studies of cognitive functions have been relatively unsuccessful in finding specific genetic variants that influence cognitive phenotypes, principally because the numbers of subjects are too small. The study of other complex phenotypes such as height suggests that many variants will be found as participant numbers increase to and beyond six figures.²⁸ Second, the cognitive GWAS consortia studies to date have used several different assessments to represent each cognitive construct across different samples, and this may have led to phenotypic heterogeneity in the derived measures.^{13,20,21,27} Third, studies to date have used samples whose genotyping has been carried out in different centres with different arrays and different quality control (QC) procedures.^{13,20,21,27} Fourth, studies have tended to examine one cognitive phenotype or domain in isolation.^{13,20,21,27}

The present study directly addresses the limitations of previous molecular genetic studies of cognitive functions. It presents genome-wide association analyses of reasoning, processing speed, declarative memory, and educational attainment in the UK Biobank sample. The number of subjects is over 100 000 for most analyses. All participants took the same cognitive tests with the same instructions. All participants included in the current analysis were of white British ancestry. Genotyping was also standardised across the same arrays and QC procedures. The study addresses three important cognitive domains and educational attainment in a single report. These advantages are likely contributors to the relative success in finding many new genetic variants associated with cognitive functions.

MATERIALS AND METHODS

This study includes baseline data from the UK Biobank Study (<http://www.ukbiobank.ac.uk>).²⁹ UK Biobank received ethical approval from the Research Ethics Committee. The REC reference for UK Biobank is 11/NW/0382. The present analyses were conducted under UK Biobank data application numbers 10279 and 7898.

Participants

The UK Biobank is a health research resource that aims to improve the prevention, diagnosis and treatment of a wide range of illnesses. Between the years 2006 and 2010, about 500 000 people aged from middle age to older age were recruited from across Great Britain. Data were collected on cognitive functions, physical and mental health, lifestyle, socio-demographic information, food intake and family medical history. For the present study, 112 151 community-dwelling individuals (58 914 females, 53 237 males) aged 40–73 years (mean = 56.91 years, s.d. = 7.93) with genome-wide genotyping were available.

Cognitive assessment

Verbal-numerical reasoning. Verbal-numerical reasoning was measured using a 13-item test presented on a touchscreen computer. The test included six verbal and seven numerical questions, all with multiple-choice answers, and had a time limit of two minutes in total. An example verbal item is: 'If Truda's mother's brother is Tim's sister's father, what relation is Truda to Tim?' (possible answers: 'aunt/sister/niece/cousin/no relation/do not know/prefer not to answer'). An example numerical item is: 'If 60 is more than half of 75, multiply 23 by 3. If not subtract 15 from 85' (possible answers: '68/69/70/71/72/do not know/prefer not to answer'). The verbal-numerical reasoning score was the total score out of 13. The Cronbach α -coefficient for the 13 items was 0.62.

Reaction time. Reaction time (RT) was measured using a computerized 'Snap' game. Participants were shown cards with symbols on a computer screen, and were directed to push a button on a nearby button box as quickly as possible with their dominant hand if the two cards had matching symbols. There were four practice trials to begin, followed by eight experimental trials, of which four had matching symbols. Each participant's RT score was their mean time (in milliseconds) to press the button for these four matching trials. The reliability (internal consistency) of these trials, using Cronbach's α , was 0.85. Before analysis, one participant with an outlying score was removed, and the data were log-transformed.

Memory. Memory was measured using a 'pairs matching' task on a touchscreen computer. Participants observed a randomly arranged grid of 12 'cards' with six pairs of matching symbols for 5 s. The symbols were then hidden, and the participant was instructed to select, from memory, the pairs that matched, in the fewest possible number of attempts. Responses were made by touching consecutive pairs on the screen. No time limit was imposed; participants were free to make as many attempts as necessary for them to correctly match all the pairs. The memory score was the total number of errors made during this task. The test was preceded by a simpler, three-pair practice version. A log +1 transformation was applied to the memory variable before analysis.

For all three cognitive tests, repeated measurements were available on a subset of participants (Table 1). The mean time difference between baseline and the repeat testing was 4.3 years (s.d. 0.9), with a range of 2.1–7.0 years. The test-retest correlations were 0.65 for verbal-numerical reasoning ($n = 4\ 696$), 0.54 for RT ($n = 20\ 188$) and 0.15 for memory ($n = 19\ 872$).

Educational attainment. To measure educational attainment, participants were asked, 'which of the following qualifications do you have? (You can select more than one)'. Possible answers were: 'college or university degree/A levels or AS levels or equivalent/O levels or GCSE or equivalent/CSEs or equivalent/NVQ or HND or HNC or equivalent/Other professional qualifications, for example, nursing, teaching/none of the above/prefer not to answer'. We

Table 1. The proportion of the phenotypic variance explained by common SNPs (h^2) and test–retest correlations for the three cognitive tests and educational attainment in UK Biobank

Phenotype	N	h^2	S.e.m.	Test–retest N	Test–retest Pearson correlation
Reaction time	94 857	0.11	0.006	20 188	0.54
Memory	95 332	0.05	0.006	19 872	0.15
Verbal–numerical reasoning	30 801	0.31	0.018	4 696	0.65
Educational attainment	94 548	0.21	0.006		

Abbreviation: SNP, single-nucleotide polymorphism.

created a binary education variable indexing whether or not each participant had attained a college or university-level degree. This follows previous studies that have used similar binary variables in GWAS studies as a successful proxy for cognitive function.²³

Genotyping and quality control

152 729 UK Biobank samples were genotyped using either the UK Biobank (N=49 979) or the UK Biobank axiom array (N=102 750). Genotyping was performed on 33 batches of ~4700 samples by Affymetrix (High Wycombe, UK). Initial QC of the genotyping data was also performed by Affymetrix. Further details are available of the sample processing specific to the UK Biobank project (<http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583>) and the Axiom array (http://media.affymetrix.com/support/downloads/manuals/axiom_2_assay_auto_workflow_user_guide.pdf). Before the release of the UK Biobank genetic data a stringent QC protocol was applied, which was performed at the Wellcome Trust Centre for Human Genetics. Details of this process can be found here (<http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580>).

Before the analyses described below, further QC measures were applied. Individuals were removed sequentially based on non-British ancestry (within those who self-identified as being British, principal component analysis was used to remove outliers, $n=32\,484$), high missingness ($n=0$), relatedness ($n=7,948$), QC failure in UK Biobank ($n=187$), and gender mismatch ($n=0$). A sample of 112 151 individuals remained for further analyses.

Imputation. An imputed data set was made available in which the UK Biobank interim release was imputed to a reference set combining the UK10K haplotype and 1000 Genomes Phase 3 reference panels. Further details can be found at the following URL: <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=157020>. The association analyses were restricted to autosomal variants with a minor allele frequency >0.1% and an imputation quality score of 0.1 or greater ($N\sim 17.3$ m SNPs).

Statistical analyses

All phenotypes were adjusted for age, gender, assessment centre, genotyping batch, genotyping array and 10 principal components to correct for population stratification before all analyses. For RT, 111 483 individuals remained for further analyses. For memory, 112 067 individuals remained for further analyses. The verbal–numerical reasoning test was added to the cognitive battery part-way through the study and was performed on a subset of 36 035 individuals who also had genotyping. For Educational Attainment, 111 114 individuals were available for further analyses.

Association analyses. Genotype–phenotype association analyses were performed on the imputed data set using SNPTest v2.5.1.³⁰ SNPTTEST v.2.5.1 can be found at the following URL: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html#introduction. An additive model was specified using the ‘frequencyist

1’ option. To account for genotype uncertainty, we analysed the genotype dosage scores.

To determine the number of independent signals from the genotype–phenotype analyses, LD clumping was used.^{31,32} The method was applied to the GWAS output for each phenotype separately, using the 1000 genomes to provide a measure of LD between the SNPs.³³ SNPs were selected for the analysis if they attained a genome-wide significant ($P < 5 \times 10^{-8}$) association with the respective phenotype. Next, SNPs in LD of $r^2 > 0.1$ and within 500 kb of the index SNP were included in the clump. SNPs from within this region were assigned to the clump if they had a P -value $< 1 \times 10^{-5}$.

Gene-based association analyses were performed using MAGMA.³⁴ The gene-based statistics were derived using the results of the GWA analyses conducted on each phenotype. Genetic variants were assigned to genes based on their position according to the NCBI 37.3 build with no additional boundary placed around the genes; this resulted in a total of 18 062 genes being analysed. The European panel of the 1000 Genomes data (phase 1, release 3) was used as a reference panel to account for linkage disequilibrium.³³ A genome-wide significance threshold for gene-based associations was calculated using the Bonferroni method ($\alpha = 0.05/18\,062$; $P < 2.8 \times 10^{-6}$).

Lookups were performed of the genome-wide significant SNP-based findings for the four UK Biobank traits in already-published GWAS of general cognitive function¹³, educational attainment²⁴ (years of education and college degree) and childhood intelligence.³⁵ For general cognitive function¹³ only a subset of the published data was available due to individual cohort restrictions on data usage ($N=36\,840$; see Supplementary Methods). We also investigated replication of published genome-wide significant findings in the present study, by comparing the SNP and gene-based association results to published findings for educational attainment (SSGAC),^{23,24} general cognitive function (CHARGE-cognitive),¹³ memory (CHARGE-cognitive),²⁰ processing speed/executive function (CHARGE-cognitive),²¹ Alzheimer’s disease (I-GAP),³⁶ intracranial volume (representing brain size)³⁷ and childhood intelligence.³⁵

Estimation of SNP-based heritability and genetic correlations. Univariate GCTA-GREML³⁸ analyses were used to estimate the proportion of variance explained by all common SNPs for each of the cognitive phenotypes and educational attainment. A relatedness cutoff of 0.025 was used in the generation of the genetic relationship matrix. LD score regression analyses were used to derive genetic correlations among the cognitive phenotypes, and between them and educational attainment. We followed the data processing pipeline devised by Bulik-Sullivan *et al.*³⁹

Polygenic prediction. Polygenic profile scores were created using PRSice⁴⁰ for the UK Biobank cognitive phenotypes and educational attainment in genotyped participants of Generation Scotland’s Scottish Family Health Study (GS, $n=19\,994$),^{18,41} and the Lothian Birth Cohort of 1936 (LBC1936, $n=1005$).^{14,42}

Individuals were removed if they had contributed to both GS and UK Biobank. Polygenic profiles are the summation of an individual's genotype across many genetic loci, weighted by the effect size estimated from a GWAS on the trait of interest. SNPs with a minor allele frequency < 0.01 were removed before creating polygenic profile scores. Genotypes were LD pruned using clumping to obtain SNPs in linkage equilibrium with an $r^2 < 0.25$ within a 200 bp window. Five polygenic profile scores were created for all four phenotypes containing SNPs according to the significance of their association with the phenotype, at P -value cutoffs of 0.01, 0.05, 0.1, 0.5 and all SNPs from the original GWAS. Linear regression models were used to examine the associations between the polygenic profiles for the UK Biobank cognitive variables and the target phenotypes in GS and LBC1936, which included multiple measures of cognitive ability, adjusted for age at measurement, sex and the first five (GS) or four (LBC1936) genetic principal components for population stratification. All models were corrected for multiple testing across all polygenic profile scores at all five thresholds in each cohort using the false discovery rate method.⁴³

Pathway and functional genomic analyses. Pathway and functional genomic analyses were performed using the GWA results for each of the cognitive phenotypes. These included DEPICT analyses⁴⁴ and reference to Regulome DB⁴⁵ (<http://www.regulome.org/>) and the Genotype-Tissue Expression Portal (<http://www.gtexportal.org>). Further details of these methods can be found in Supplementary Methods.

RESULTS

A description of the UK Biobank cohort is presented in Supplementary Table S1. Just under one-third of the sample ($n = 33\,852$, 30.5%) had a college or university degree. When scored such that higher scores represented better performance, the phenotypic correlations between the cognitive tests were all positive (Table 2). Verbal-numerical reasoning correlated with RT and memory at $r = 0.16$ and 0.18 , respectively. RT correlated with memory at $r = 0.12$. The point-biserial correlations of educational attainment with verbal-numerical reasoning, RT, and memory were $r = 0.34$, 0.10 and 0.05 , respectively.

The results of the GWAS analyses are presented below; Manhattan and QQ Plots for each trait are shown in Figure 1.

Verbal-numerical reasoning

A total of 149 SNPs from three genomic regions were significantly associated with verbal-numerical reasoning scores (Supplementary Table S2). Three independent signals were identified. The strongest signal was on chromosome 22 in a region that includes:

the cytochrome P450 gene *CYP2D6*, which is linked to drug metabolism; NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14 kDa (*NDUFA6*), which is involved in mitochondrial function;⁴⁶ and septin 3 (*SEPT3*), which is associated with Alzheimer's disease.⁴⁷ The two other regions included: SNPs in phosphodiesterase 1C, calmodulin-dependent 70 kDa (*PDE1C*), a calmodulin-dependent PDE that is stimulated by a calcium-calmodulin complex;⁴⁸ and a single SNP in Fucosyltransferase 8 (alpha (1,6) fucosyltransferase (*FUT8*), which catalyses the transfer of fucose from GDP-fucose to N-linked type complex glycopeptides.⁴⁹ These were on chromosomes 7 and 14, respectively.

The gene-based analysis of verbal-numerical reasoning identified 17 significant genes from across seven regions, including multiple hits on chromosome 22, such as *SEPT3*, *CYP2D6* and *NDUFA6* (Supplementary Table S3). Other gene-based hits linked to neurobiological pathways include: Ataxin2-like (*ATXN2L*) on chromosome 16 (a member of the spinocerebellar ataxia family which is associated with neurodegenerative disorders); amyloid beta (A4) Precursor Protein-Binding, Family A, Member 1 (*APBA1*) on chromosome 9, which interacts with the Alzheimer's disease amyloid precursor protein;⁵⁰ and SH2B Adaptor Protein 1 (*SH2B1*) on chromosome 16, previously associated with type 2 diabetes.⁵¹

The proportion of variance in verbal-numerical reasoning that was explained by all common genetic variants was 31% (GCTA-GREML estimate 0.31, s.e.m. = 0.018) (Table 1).

Reaction time

For RT there were 36 SNP hits at the genome-wide significance threshold spanning two regions, one on chromosome 2 and the other on chromosome 12 (Supplementary Table S2). Two independent signals were identified from these top hits, including: a variant in the SH2B Adaptor Protein 3 (*SH2B3*) gene on chromosome 12, which is involved in signalling activities by growth factor and cytokine receptors; and a variant in spermatogenesis-associated, serine-rich 2-like (*SPATS2L*) (on chromosome 2). The chromosome 12 locus has previously been linked to a wide spectrum of disease susceptibilities affecting multiple organs, including neurodegenerative conditions and longevity.⁵²

In the gene-based analysis, 23 genes from across nine regions were identified as having significant associations with RT (Supplementary Table S3). These included: *SH2B3* and Ataxin2 (*ATXN2*), associated with spinocerebellar ataxia 2⁵³ on chromosome 12; autophagy/beclin-1 regulator 1 (*AMBRA1*), important in autophagy and the development of the nervous system⁵⁴ on chromosome 11; diacylglycerol kinase, zeta (*DGKZ*), involved in intracellular signalling,⁵⁵ on chromosome 5; and neuron navigator 1 (*NAV1*), expressed in the nervous system and thought to have a role in neuronal development and regeneration,⁵⁶ on chromosome 1.

Table 2. Descriptive statistics and phenotypic (below diagonal) and genetic (above diagonal) correlations for the UK Biobank cognitive and educational variables in all genotyped participants

Variable	Mean (s.d.)	Genetic (above)/phenotypic (below)/correlations			
		Reaction time	Memory	Verbal-numerical reasoning	Educational attainment
Reaction time (ms)	555.08 (112.69)	—	0.179 (0.06)*	0.206 (0.05)*	0.066 (0.04)
Memory (errors)	4.06 (3.23)	0.116 (0.003)*	—	0.437 (0.06)*	0.126 (0.05) [†]
Verbal-numerical reasoning (maximum score 13)	6.16 (2.10)	0.156 (0.005)*	0.176 (0.005)*	—	0.729 (0.04)*
Educational attainment	30.5% With degree	0.099 (0.003)*	0.052 (0.003)*	0.338 (0.005)*	—

Genetic correlations are based on the results of genome-wide association studies of the UK Biobank variables. S.e.m. for the correlations are shown in parentheses. For the phenotypic variables, Pearson correlations were used for continuous-continuous correlations and point-biserial correlations for continuous-categorical correlations. All variables are coded such that higher scores indicate better performance. *indicates P -value < 0.0001; [†]indicates P -value < 0.05. This table has been published previously.⁶²

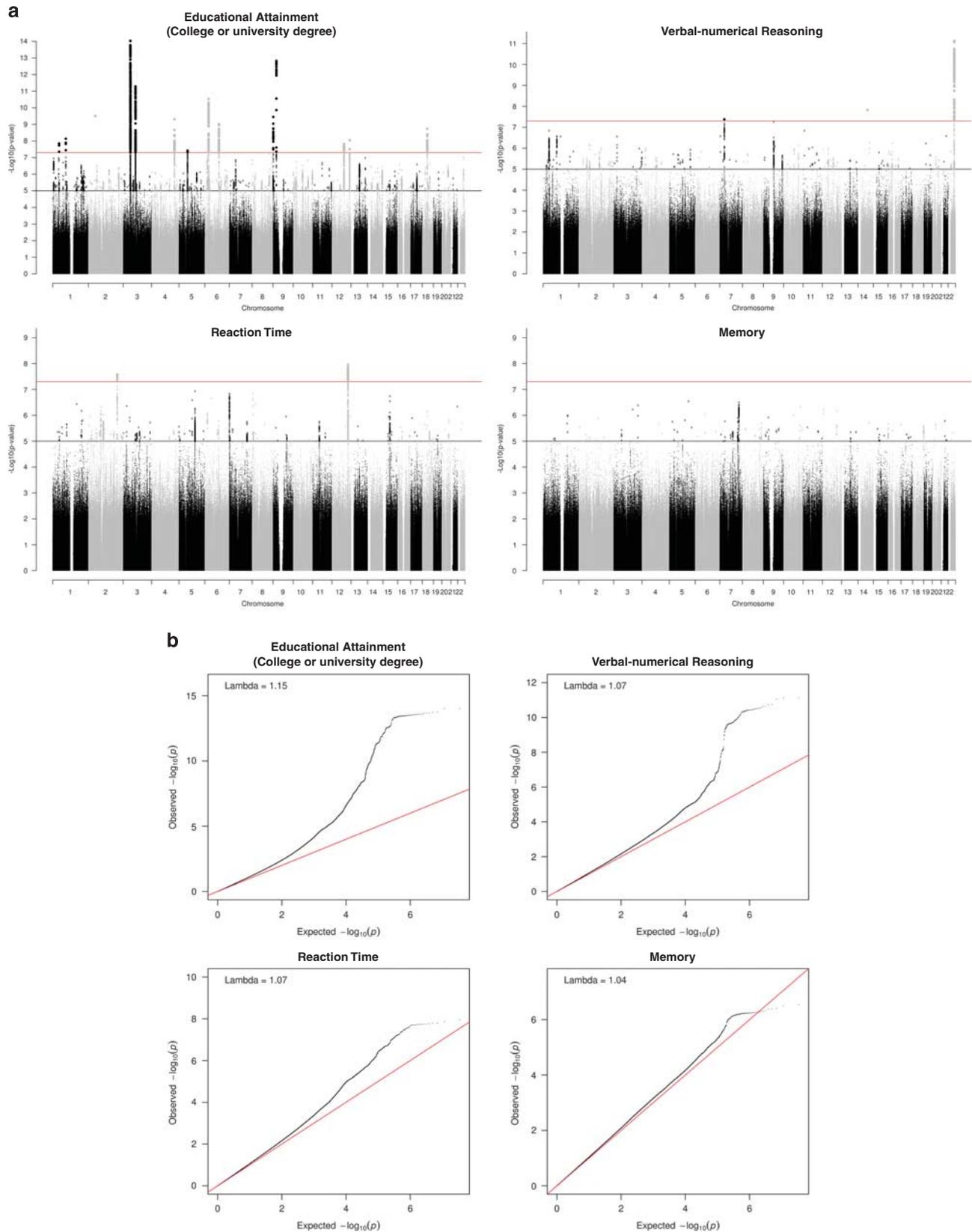


Figure 1. (a) Manhattan and (b) Q-Q plots of P -values of the SNP-based association analyses for each of the cognitive phenotypes: educational attainment, verbal-numerical reasoning, reaction time and memory. The red line indicates the threshold for genome-wide significance ($P < 5 \times 10^{-8}$); the grey line indicates the threshold for suggestive significance ($P < 1 \times 10^{-5}$). SNP, single-nucleotide polymorphism.

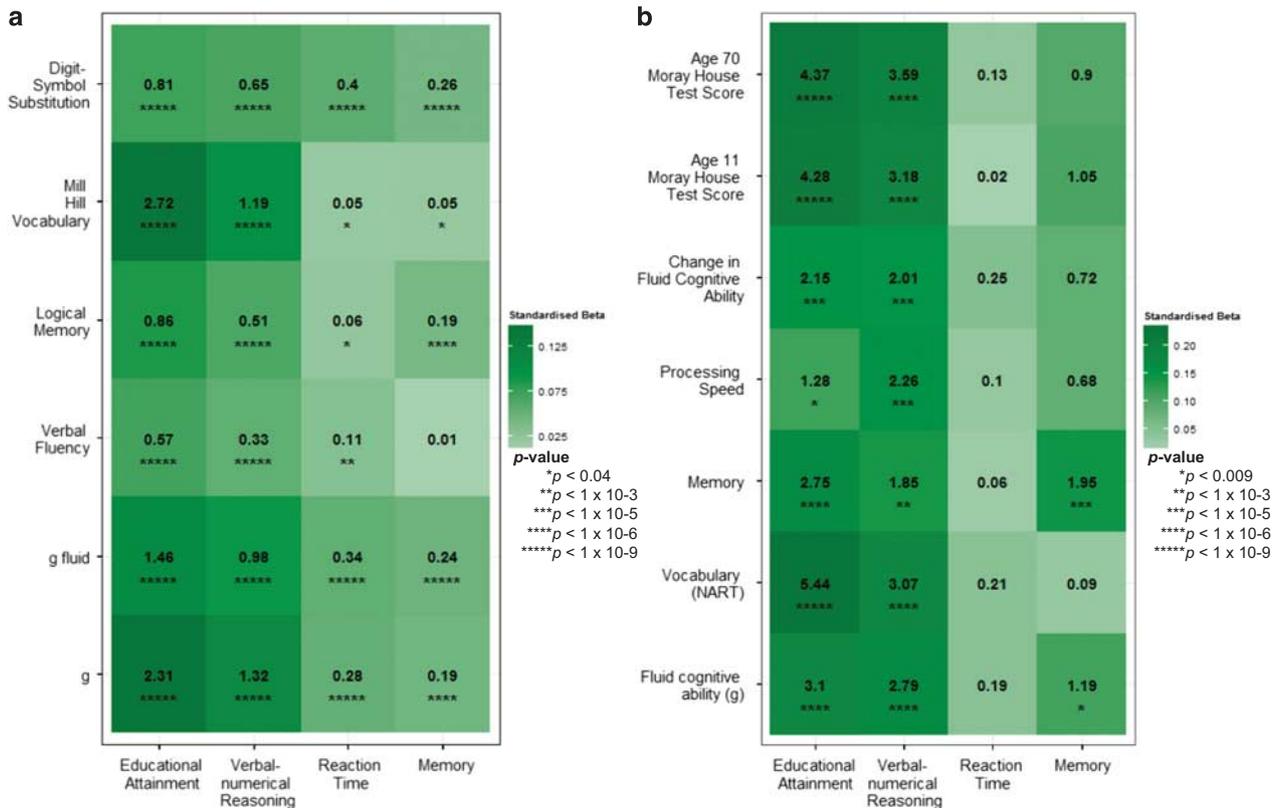


Figure 2. Heat maps of associations between the polygenic profile scores for cognitive phenotypes in UK Biobank and cognitive ability in Generation Scotland (a) and the Lothian Birth Cohort 1936 (b). Stronger associations are indicated by darker shades. The amount of variance (%) explained is indicated for each association. Further information can be found in Supplementary Table 4.

The GCTA-GREML estimate showed that 11% of the variance in RT can be explained by common genetic variants (estimate 0.11, s.e.m. = 0.006) (Table 1).

Memory

There were no genome-wide significant SNP-based findings for memory scores, despite the sample size being as large as for both RT and educational attainment and three times larger than that available for verbal–numerical reasoning. Two gene-based results from separate regions were identified for memory (Supplementary Table S3). These were: the exocyst complex component 4 (*EXOC4*), a component of the exocyst complex which is required for targeting exocytic vesicles to specific docking sites on the plasma membrane and has been associated with rate of cognitive decline in Alzheimer’s disease;⁵⁷ and exostosin glycosyltransferase 1 (*EXT1*), associated with type I multiple exostoses,⁵⁸ on chromosomes 7 and 8 respectively.

The SNP-based GCTA-GREML estimate was 0.05 (s.e.m. = 0.006), that is, 5% of the variance in memory test scores can be explained by common genetic variants (Table 1).

Educational attainment

There were 1 115 SNPs that were associated with educational attainment at the genome-wide significance threshold ($P < 5 \times 10^{-8}$) (Supplementary Table S2). The most significant was on chromosome 3 in a gene-dense region with more than 30 different genes. The top hit, in the Ras homologue family member A (*RHOA*) gene is an insertion/deletion and therefore was not included in the LD clumping analysis. After LD clumping, there were 15 independent signals, including: four on chromosome 3, two of which were in the cell adhesion molecule 2 (*CADM2*) gene,

which is involved in synapse organization;⁵⁹ and one in the CaM kinase-like vesicle-associated (*CAMKV*) gene.

A gene-based analysis identified 95 genes from across 28 regions that were significantly associated with Educational Attainment (Supplementary Table S3). The top hit from the gene-based analysis was MON1 secretory trafficking family member A (*MON1A*), involved in membrane trafficking,⁶⁰ which is on chromosome 3. The *CADM2* gene was also significant. *ATXN2L* and *SH2B1*, which were significant in the verbal–numerical reasoning gene-based analysis, were also significantly associated with educational attainment.

The common genetic variants from the genotyped SNPs explained 21% of the variance in educational attainment (GCTA-GREML estimate 0.21, s.e.m. = 0.006) (Table 1).

Polygenic profile scoring

The GWAS results for the three cognitive tests and the educational attainment measure were used to build polygenic profile scores in two independent cohorts, the Lothian Birth Cohort 1936 (LBC1936) and GS. Significant predictions were observed across almost all thresholds for all of the predictors for the cognitive phenotypes measured in LBC1936 and GS (Figure 2 and Supplementary Table S4). The largest proportion of variance explained in LBC1936 was 5.4% for a vocabulary-based test (the National Adult Reading Test)⁶¹ using the Educational Attainment polygenic score with a SNP inclusion threshold of all SNPs from the GWAS. In GS the best prediction was also for the vocabulary measure; 2.8% of the variance was explained, again by the educational attainment polygenic score, and with a SNP inclusion threshold of $P < 0.50$.

Genetic correlations

LD score regression was used to estimate the genetic correlations between the three cognitive traits and educational attainment. When scored such that higher scores represented better performance, the genetic correlations were all positive (Table 2). The largest genetic correlations were observed between verbal–numerical reasoning and educational attainment ($r_g=0.73$, s.e.m.=0.04), memory ($r_g=0.44$, s.e.m.=0.06) and RT ($r_g=0.21$, s.e.m.=0.05). These correlations have been published previously.⁶²

Replication of SNP- and gene-based hits

We sought replication of the top SNP-based findings in published GWAS of general cognitive function,¹³ educational attainment²⁴ (years of education and college degree), and childhood cognitive function³⁵ (Supplementary Table S5). We also compared SNP- and gene-based hits (where available) from the previous literature with the corresponding results from the present UK Biobank analysis (Supplementary Tables S6 and S7).

We first describe the SNP-based lookup in the already-published GWAS (Supplementary Table S5). We present lookups for all available SNPs, as not all SNPs from the current study were available due to differences in imputation reference panels. We find that, of the 1115 genome-wide significant SNPs associated with educational attainment, 327 (general cognitive function), 326 (years of education), 326 (college degree) and 267 (childhood intelligence) were available in the published GWAS. Of these 158, 240, 211 and 47, respectively, showed replication at $P < 0.05$, for general cognitive function, years of education, college degree and childhood cognitive function. For verbal–numerical reasoning we report 149 SNPs reaching genome-wide significance; 37 were available for general cognitive function, years of education and college degree, and 29 for childhood intelligence. Thirty-six SNPs in the chromosome 22 region replicate for general cognitive function, one SNP on chromosome 7 demonstrated replication with childhood intelligence, and no replication was observed with either years of education or college degree. None of the 36 SNPs associated with RT replicated in any of the above-listed published GWAS.

The SNP-based lookups, within this new UK Biobank sample, of previously reported genome-wide significant findings are detailed in Supplementary Table S6. Replication is reported where $P < 0.05$ in the UK Biobank GWASs. For the 13 genome-wide significant SNPs in the 2015 GWAS of general cognitive function,¹³ 11 SNPs in the chromosome 6 region were replicated for educational attainment, and verbal–numerical reasoning. A single SNP on chromosome 14 (rs17522122 in the *AKAP6* gene) was replicated in verbal–numerical reasoning, RT, and memory. Of the three memory hits reported by DeBette *et al.*,²⁰ one SNP (rs4420638 in the *APOC1* gene) replicated in the UK Biobank memory GWAS.

Of the three educational attainment SNP hits (two for college degree, rs11584700, rs4851266; one for years of education, rs9320913) reported by Rietveld *et al.*,²⁴ all three replicated in UK Biobank for educational attainment, and two SNPs, rs4851266 and rs9320913, also replicated for verbal–numerical reasoning. Of the three SNPs associated with educational attainment and cognitive function (rs1487441, rs7923609 and rs2721173) in Rietveld *et al.*,²³ all demonstrated replication with educational attainment in UK Biobank, rs1487441 and rs2721173 with verbal–numerical reasoning, and rs2721173 with RT.

Of the 21 genome-wide significant hits reported in the most recent Alzheimer's disease GWAS, one SNP, rs983392, showed replication with memory in UK Biobank.

One SNP, rs17689882, in the *CRHR1* gene that was associated with intracranial volume³⁷ was replicated in UK Biobank with both educational attainment and verbal–numerical reasoning.

For the gene-based lookups, the *HMGN1* gene reported in the general cognitive function paper¹³ did not replicate for any of the traits (Supplementary Table S7). The *FNBP1L* gene, which has been associated with childhood cognitive ability,³⁵ was replicated in the educational attainment analysis ($P=0.004$). Of the seven gene-based hits for educational attainment²⁴ (college versus no college), two (*PIK3C2B* and *TET2*) were replicated in UK Biobank for the educational attainment variable. None of the genes was replicated for the other three traits. For the years of education variable examined by Rietveld *et al.*,²⁴ 12 out of the 16 significant genes were replicated in the UK Biobank educational attainment analysis. Five of these genes also replicated for verbal–numerical reasoning. There was no overlap for the RT or memory traits. Of the genes linked to educational attainment and cognitive function in Rietveld *et al.*,²³ *NRXN1* was associated with verbal–numerical reasoning and *POU3F2* with memory in UK Biobank.

DEPICT results

For the educational attainment phenotype, gene prioritization as implemented in DEPICT indicated a role for 28 genes at false discovery rate < 0.05 (Supplementary Table S8). No genes showed statistically significant links to verbal–numerical reasoning, RT or memory. In the gene-set analyses, two gene sets were significantly enriched for verbal–numerical reasoning; these were 'regulation of cell morphogenesis' (GO:0022604), false discovery rate ≤ 0.01 , and 'CLTC PPI subnetwork' (ENSG00000141367), false discovery rate ≤ 0.05 . No significant results were found with educational attainment, RT or memory.

The tissue enrichment analyses yielded no significant findings for any of the cognitive function phenotypes. However, nominally significant results for the educational attainment, memory, and RT phenotypes showed enrichment of central nervous system tissue types.

Functional analysis and gene expression

Using the GTEx database (<http://www.broadinstitute.org/gtex/>), three cis-eQTL associations were identified for the 18 independent genome-wide significant SNPs found in the present report that were also included in this database (Supplementary Table S9). These were rs13086611, rs11130222 (an intronic SNP in *CAMKV*), and rs2142694 and they potentially regulate *FAM212A*, *RBM6* and *CYP2D6/SMDT1/NAGA*, respectively. For this study, data mining of regulatory elements was restricted to normal cell lines/tissues. There was evidence of regulatory elements associated with all 12 of the independent genome-wide significant SNPs included in the Regulome DB database. (<http://www.regulomedb.org/>) (Supplementary Table S9).

DISCUSSION

The results of the present study make novel contributions to three scientific aims of GWAS: helping towards identifying specific mechanisms of genomic variation; describing the genetic architecture of complex traits; and predicting phenotypic variation in independent samples. The most important novel contribution of the present study is the discovery of many new genome-wide significant genetic variants associated with reasoning ability, cognitive processing speed and the attainment of a college or university degree. The study provided robust estimates of the SNP-based heritability of the four cognitive variables and their genetic correlations. The study makes important steps toward genetic consilience, because several of the genomic regions identified by the present analyses have previously been associated in GWASs of general cognitive function, executive function, educational attainment, intracranial volume, neurodegenerative disorders and Alzheimer's disease. The study was successful in

using the GWAS results from UK Biobank to predict cognitive variation in new samples.

The present study identified some novel genes/loci associated with individual differences in cognitive functions. These include, for RT, a locus on chromosome 12q24 containing the growth repressor *SH2B3*, and the RNA processing factor *ATXN2*, previously associated with a large number of disease susceptibilities including: type 1 diabetes, multiple sclerosis, spinocerebellar ataxia type 2, Parkinson's disease, longevity and many more.⁵² Phenotypically, processing speed and longevity are associated.⁶³ Therefore, it is possible that this locus is influencing RT through its influence on a number of health traits. The most significant finding for verbal–numerical reasoning was for a gene-dense region on chromosome 22, and replication of this region was observed with general cognitive function. This region contains the cytochrome P450 gene (*CYP2D6*), which uses hydroxytryptamines (such as serotonin) and neurosteroids as endogenous substrates⁶⁴ and may explain some of the links between cognitive functions and mental illnesses. Given the associations of this region with drug metabolism, it would be interesting to investigate this finding further with specific reference to medication use and psychiatric disease history. Brain size and cognitive ability are correlated phenotypically,⁶⁵ and the significant hit in the GWAS of intracranial volume³⁷ was also significant in SNP- and gene-based analyses in the present study.

Previously published associations between a number of genes and cognitive function and educational attainment have been replicated. *CADM2* (chromosome 3), which was previously associated with executive functioning and processing speed²¹ and showed suggestive association with general cognitive function,¹³ was associated with educational attainment in this study. *CADM2* encodes a synaptic cell adhesion molecule and is important in maintaining synaptic circuitry of the central nervous system.⁶⁶ The two genes linked previously to educational attainment and cognitive function²³ and that were associated with verbal–numerical reasoning and memory in UK Biobank are associated with the regulation of post-synaptic *N*-methyl *D*-aspartate receptors (*NRXN1*) and neuronal differentiation (*POU3F2*). Genetic variation in the *N*-methyl *D*-aspartate receptor complex has been previously shown to have an enriched association with cognitive abilities.⁶⁷ Several specific SNPs previously associated with general cognitive function¹³ were associated with educational attainment and verbal–numerical reasoning, with single hits associated with RT and memory phenotypes in UK Biobank. This suggests that educational attainment can be used to some extent as a proxy for cognitive function (particularly verbal–numerical reasoning) in this sample. The genes *ATXN2L* and *SH2B1* (both on chromosome 16), previously linked to duration of education,²⁴ were associated with both educational attainment and verbal–numerical reasoning in UK Biobank, as were three SNPs identified by Rietveld *et al.*²³ as being associated with education and cognitive function.

Evidence of regulatory elements associated with all 12 of the genome-wide significant independent SNPs included in the Regulome DB database was identified within normal tissues and cell lines. The regulatory elements identified include position weight matrix, transcription factor binding sites, histone modifications, DNase hypersensitive sites and FAIRE sites. This evidence suggests that the associated SNPs are in sites of active transcription and could have a regulatory role on transcription. Three of the SNPs are potentially eQTLs.

The SNP-based estimate of heritability for verbal–numerical reasoning (31%) was highly consistent with previous estimates based on a general cognitive ability phenotype that had been composed using three or more diverse cognitive tests.^{13,14,18} Using the summary GWAS data from the present study to predict cognitive variation in independent samples (Supplementary Table S4) produced the largest R^2 values in this field to date, with sometimes over 5% of variance explained, especially in the more

crystallized cognitive functions such as vocabulary. Previously, values of 1 to 2% have been typical.¹³

The present study has several strengths. It has the largest sample size to date for a GWAS of any cognitive phenotype. It has the added advantage of being a single sample rather than a patchwork of small GWASs requiring meta-analysis. The tests offered were the same for all UK Biobank participants, so the phenotypic heterogeneity of previous cognitive GWASs has been overcome. It has phenotypes covering three important cognitive domains and a measure of educational attainment that could be used as a proxy for cognitive function. All individuals were of white British ethnicity, which minimized population stratification. The genotyping and its QC were performed in a consistent way.

The pattern of the present study's results, whereby the educational attainment and verbal–numerical reasoning variables had higher heritability and more genome-wide significant hits than RT or memory, can be understood with reference to some regularities in the relevant literature and, of course, on time constraints placed on data collection in a study that has a wide-ranging health remit. First, general cognitive ability, or strong indicators of it, tend to be more heritable than specific cognitive functions such as processing speed and memory.^{8,10,12,68} Second, tests of verbal ability and reasoning are among those tests that have higher loadings on the latent trait of general cognitive ability, and tests of memory and processing speed have lower loadings.^{15,69–71} Third, the RT and memory tests in UK Biobank were handicapped further by being very brief. The RT test included a far smaller number of trials than is typical for large surveys in the UK, which have used 40 trials in choice RT procedures.^{72,73} The memory test was based on the recall of a single 12-item matrix with six pairs of stimuli. This is both a brief and unusual type of test in the field of declarative memory; more is known about the psychometric characteristics and genetic foundations of declarative memory tests such as word list and paragraph recall.²⁰ The test–retest correlation of the memory variable was particularly low ($r=0.15$). When compared with phenotypic correlations of similar tests in the Lothian Birth Cohort 1936, correlations for the UK Biobank tests are about or more than 0.1 lower (Supplementary Table S10). Fourth, recent evidence from previous large studies make the strong evidence of a genetic contribution to educational attainment in the UK Biobank sample unsurprising. Despite their often being used as social background/environmental variables,⁶⁸ educational attainments are substantially heritable,⁷⁴ with many of the same genes affecting different academic subjects,⁷⁵ and they have high phenotypic and genetic correlations with cognitive ability test scores, especially general cognitive ability.^{25,68,76} Moreover, it has already been demonstrated that educational attainment is useful as a proxy variable for cognitive ability in GWAS analyses.²³

This accumulating evidence is consistent with the interpretation that, to some extent, educational attainments are a product of genetic contributions to cognitive ability, but with two emphatic qualifications. First, it is obvious that there are other—especially social—determinants of whether or not people achieve certain educational outcomes.⁶⁸ Second, there is evidence that the variation in educational attainments that is caused by genetic differences is shared with traits other than intelligence, such as personality dimensions.^{75,77} Therefore, we predict that not all of the genome-wide significant hits associated with the attainment of a college or university degree in the present study will be associated with cognitive differences; some might be associated with personality and other heritable, educational relevant traits.

The present study has limitations. Although the sample size is large for its field, GWASs of other complex traits such as height demonstrate that even larger sample sizes are required, and that these will reveal many more significant genetic variants. Another limitation concerns the cognitive tests. The three measures used were non-standard, bespoke tests. They did show the expected

positive correlations with one another (and with education), but these correlations were lower than is typically found in studies of cognitive batteries which have tests with more items. The limitations of the RT and memory tests were discussed above. The verbal-numerical reasoning test had only 13 items, some of which had floor effects, and it had modest internal consistency. With more in-depth, albeit time-demanding tests, we expect to find even more genetic variants linked to differences in cognitive ability.⁷⁸ Indeed, the educational variable heralds this. The participants in the sample were all white British, which restricts the generalizability of the results, which require extending to other groups. We also note that there are likely to be other types of genetic variation contributing to cognitive differences in addition into the common SNP variations studied here.

When genetic data on the full ~500 000 UK Biobank sample becomes available it is certain that more genetic variants will be found that are relevant for cognitive variation. This would enable robust replication of our current results in addition to other useful analyses. For example, sex- and age-moderation of genetic determinants could be studied with high power. It is possible, for example, that, as we have found for *APOE* variation,¹³ some genetic variants will have stronger effects at some ages than others, or even no effects whatsoever at some ages. There are also plans to enhance the cognitive testing in UK Biobank, which will afford the genetic study of additional and more detailed cognitive tests.

CONCLUSION

The present results make advances in, and encourage much further work on the genetic foundations of cognitive differences. Until recently, GWASs of cognitive functions had provided converging information about their polygenic architecture—especially via SNP-based heritability estimates—and modest power to predict cognitive phenotypes in independent studies. However, they were near-bereft of significant, replicable genetic variants that could be followed-up to understand why some people have more efficient cognitive functioning than others. This drought is ending; work can begin on the genetically driven biological mechanisms of cognitive differences and the biological foundations of the many associations between cognitive abilities and bio-medical, health and social variables.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This research was conducted, using the UK Biobank Resource, in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. GS has received core funding from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). We are grateful to the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole GS team. Genotyping was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility (WTCRF), Edinburgh, Scotland, funded by the MRC. The Quantitative Trait Locus team at the Human Genetics Unit are funded by the MRC. Data collection from the Lothian Birth Cohort 1936 (LBC1936) was supported by the Disconnected Mind project, funded by Age UK. We thank the LBC1936 members for their participation, and the whole LBC1936 team for their work on the phenotypes. BBSRC provided funding for the genetic testing in LBC1936 (BB/F019394/1). Genotyping was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility (WTCRF), Edinburgh, Scotland, funded by the MRC. We thank the Scottish Council for Research in Education Centre, University of Glasgow, for access to the results from the Scottish Mental Survey 1947. We are grateful for the use of summary data from a subset of cohorts included in the CHARGE consortium cognitive working group general cognitive function GWAS (see Supplementary Methods).

REFERENCES

- Deary IJ, Weiss A, Batty GD. Intelligence and personality as predictors of illness and death: how researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychol Sci Public Interest* 2010; **11**: 53–79.
- Gale CR, Deary IJ, Boyle SH, Barefoot J, Mortensen LH, Batty GD. Cognitive ability in early adulthood and risk of five specific psychiatric disorders in mid life: the Vietnam Experience Study. *Arch Gen Psychiatr* 2008; **65**: 1410–1418.
- Eggermont LH, de Boer K, Muller M, Jaschke AC, Kamp O, Scherder EJ. Cardiac disease and cognitive impairment: a systematic review. *Heart* 2012; **98**: 1334–1340.
- Hedman AM, van Haren NE, van Baal CG, Kahn RS, Hulshoff Pol HE. IQ change over time in schizophrenia and healthy individuals: a meta-analysis. *Schizophr Res* 2013; **146**: 201–208.
- Strenze T. Intelligence and socioeconomic success: a meta-analytic review of longitudinal research. *Intelligence* 2007; **35**: 401–426.
- Salthouse TA. Localizing age-related individual differences in a hierarchical structure. *Intelligence* 2004; **32**: 541–561.
- Wilson RS, Beckett LA, Barnes LL, Scheider JA, Bach J, Evans DA *et al*. Individual differences in rates of change in cognitive abilities of older persons. *Psychol Aging* 2002; 179–193.
- Bouchard TJ. Genes, evolution and intelligence. *Behav Genet* 2014; **44**: 549–577.
- Deary IJ, Johnson W, Houlihan LM. Genetic foundations of human intelligence. *Hum Genet* 2009; **126**: 215–232.
- Deary IJ, Penke L, Johnson W. The neuroscience of human intelligence differences. *Nat Rev Neurosci* 2010; **11**: 201–211.
- Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM *et al*. The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol Psychiatr* 2010; **15**: 1112–1120.
- Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. *Mol Psychiatr* 2015; **20**: 98–108.
- Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S *et al*. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53,949). *Mol Psychiatr* 2015; **20**: 183–192.
- Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D *et al*. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatr* 2011; **16**: 996–1005.
- Carroll JB. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press: Cambridge, UK, 1993.
- Deary IJ. Intelligence. *Curr Biol* 2013; **23**: R673–R676.
- Deary IJ, Spinath FM, Bates TC. Genetics of intelligence. *Eur J Hum Genet* 2006; **14**: 690–700.
- Marioni RE, Davies G, Hayward C, Liewald DC, Kerr SM, Campbell A *et al*. Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 2014; **44**: 26–32.
- Trzaskowski M, Shakeshaft NG, Plomin R. Intelligence indexes generalist genes for cognitive abilities. *Intelligence* 2013; **41**: 560–655.
- Debette S, Ibrahim-Verbaas CA, Bressler J, Schuur M, Smith A, Bis JC *et al*. Genome-wide studies of verbal declarative memory in non-demented older people: the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium. *Biol Psychiatry* 2015; **77**: 749–763.
- Ibrahim-Verbaas CA, Bressler J, Debette S, Schuur M, Smith AV, Bis JC *et al*. GWAS for executive function and processing speed suggests involvement of the *CADM2* gene. *Mol Psychiatr* 2015; **21**: 189–197.
- Rietveld CA, Conley D, Eriksson N, Esko T, Medland SE, Vinkhuyzen AA *et al*. Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychol Sci* 2014; **25**: 1975–1986.
- 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. *Proc Natl Acad Sci USA* 2014; **111**: 13790–13794.
- Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW *et al*. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013; **340**: 1467.
- Calvin CM, Deary IJ, Webbink D, Smith P, Fernandes C, Lee SH *et al*. Multivariate genetic analyses of cognition and education from two population samples of 174,000 and 166,000 school children. *Behav Genet* 2012; **42**: 699–710.
- Krapohl E, Plomin R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatr* 2015; **21**: 437–443.
- Trampush JW, Lencz T, Knowles E, Davies G, Guha S, Pe'er I *et al*. Independent evidence for an association between general cognitive ability and a genetic locus for educational attainment. *Am J Med Genet B: Neuropsychiatr Genet* 2015; **168**: 363–373.
- Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S *et al*. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014; **46**: 1173–1186.

- 29 Collins R. What makes UK Biobank special? *Lancet* 2012; **379**: 1173–1174.
- 30 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies via imputation of genotypes. *Nat Genet* 2007; **39**: 906–913.
- 31 Purcell S. PLINK v1.07. <http://pngu.mgh.harvard.edu/purcell/plink/>.
- 32 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al*. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 559–575.
- 33 The 1000 Genomes Project. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.
- 34 deLeeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS Data. *PLoS Comput Biol* 2015; **11**: 4.
- 35 Benyamin B St, Pourcain B, Davis OS, Davies G, Hansell NK, Brion M-JA *et al*. Childhood intelligence is heritable, highly polygenic and associated with *FNBP1L*. *Mol Psychiatr* 2014; **19**: 253–258.
- 36 Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al*. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**: 1452–1458.
- 37 Hibar D, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N *et al*. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**: 224–229.
- 38 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al*. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**: 565–569.
- 39 Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**: 291–295.
- 40 Eusden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics* 2015; **31**: 1466–1468.
- 41 Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ *et al*. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* 2006; **7**: 74.
- 42 Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V *et al*. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* 2007; **7**: 28.
- 43 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodological)* 1995; 289–300.
- 44 Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J *et al*. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 2015; **6**: 5890.
- 45 Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–1797.
- 46 Emahazion T, Brookes AJ. Mapping of the NDUFA2, NDUFA6, NDUFA7, NDUFB8, and NDUFS8 electron transport chain genes by intron based radiation hybrid mapping. *Cytogenet Cell Genet* 1998; **82**: 114.
- 47 Takehashi M, Alioto T, Stedeford T, Persad AS, Banasik M, Masliah E *et al*. Septin 3 gene polymorphism in Alzheimer's disease. *Gene Expr* 2004; **11**: 263–270.
- 48 Repaske DR, Swinnen JV, Jin SL, Van Wyk JJ, Conti M. A polymerase chain reaction strategy to identify and clone cyclic nucleotide phosphodiesterase cDNAs. Molecular cloning of the cDNA encoding the 63-kDa calmodulin-dependent phosphodiesterase. *J Biol Chem* 1992; **267**: 18683–18688.
- 49 Miyoshi E, Noda K, Yamaguchi Y, Inoue S, Ikeda Y, Wang W *et al*. The alpha1-6-fucosyltransferase gene and its biological significance. *Biochim Biophys Acta* 1999; **1473**: 9–20.
- 50 Sullivan SE, Dillon GM, Sullivan JM, Ho A. Mint proteins are required for synaptic activity-dependent amyloid precursor protein (APP) trafficking and amyloid β generation. *J Biol Chem* 2014; **289**: 15374–15383.
- 51 Rui L. SH2B1 regulation of energy balance, body weight, and glucose metabolism. *World J Diabetes* 2014; **5**: 511–526.
- 52 Auburger G, Gispert S, Lahut S, Omür O, Damrath E, Heck M *et al*. 12q24 locus association with type 1 diabetes: SH2B3 or ATXN2? *World J Diabetes* 2014; **5**: 316–327.
- 53 Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H *et al*. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 1996; **14**: 277–284.
- 54 Fimia GM, Stoykova A, Romagnoli A, Giunta L, Di Bartolomeo S, Nardacci R *et al*. Ambra1 regulates autophagy and development of the nervous system. *Nature* 2007; **447**: 1121–1125.
- 55 Sakane F, Kanoh H. Molecules in focus: diacylglycerol kinase. *Int J Biochem Cell Biol* 1997; **29**: 1139–1143.
- 56 Maes T, Barceló A, Buesa C. Neuron navigator: a human gene family with homology to unc-53, a cell guidance gene from *Caenorhabditis elegans*. *Genomics* 2002; **80**: 21–30.
- 57 Sherva R, Tripodis Y, Bennett DA, Chibnik LB, Crane PK, de Jager PL *et al*. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement* 2014; **10**: 45–52.
- 58 Hecht JT, Hogue D, Wang Y, Blanton SH, Wagner M, Strong LC *et al*. Hereditary multiple exostoses (EXT): mutational studies of familial EXT1 cases and EXT-associated malignancies. *Am J Hum Genet* 1997; **60**: 80–86.
- 59 Frei JA, Andermatt I, Gesemann M, Stoeckli ET. The SynCAM synaptic cell adhesion molecules are involved in sensory axon pathfinding by regulating axon-axon contacts. *J Cell Sci* 2014; **127**: 5288–52302.
- 60 Bagley DC, Paradar PN, Kaplan J, Ward DM. Mon1a protein acts in trafficking through the secretory apparatus. *J Biol Chem* 2012; **287**: 25577–25588.
- 61 Nelson HE, Willison J. *National Adult Reading Test (NART)*. nfer-Nelson : Windsor, UK, 1991.
- 62 Hagenars SP, Harris SE, Davies G, Hill WD, Liewald DCM, Ritchie SJ *et al*. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112 151) and 24 GWAS consortia. *Mol Psychiatr*. e-pub ahead of print.
- 63 Deary IJ, Der G. Reaction time explains IQ's association with death. *Psychol Sci* 2005; **16**: 64–69.
- 64 Wang B, Yang LP, Zhang XZ, Huang SQ, Bartlam M, Zhou SF. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. *Drug Metab Rev* 2009; **41**: 573–643.
- 65 Ritchie SJ, Booth T, Valdes Hernandez MDC, Corley J, Munoz Maniega S, Gow AJ *et al*. Beyond a bigger brain: multivariable brain imaging and intelligence. *Intelligence* 2015; **51**: 47–56.
- 66 Thomas LA, Akins MR, Biederer T. Expression and adhesion profiles of SynCAM molecules indicate distinct neuronal functions. *J Comp Neurol* 2008; **510**: 47–67.
- 67 Hill WD, Davies G, Van De Lagemaat LN, Christoforou A, Marioni RE, Fernandes CPD *et al*. Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl Psychiatry* 2014; **4**: e341.
- 68 Deary IJ, Johnson W. Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *Int J Epidemiol* 2010; **39**: 1362–1369.
- 69 Sternberg RJ. Component processes in analogical reasoning. *Psychol Rev* 1977; **81**: 353–378.
- 70 Johnson W, Bouchard TJ, Krueger RF, McGue M, Gottesman IA. Just one *g*: consistent results from three test batteries. *Intelligence* 2004; **32**: 95–107.
- 71 Johnson W, Deary IJ. Placing inspection time, reaction time, and perceptual speed in the broader context of cognitive ability: the VPR model in the Lothian Birth Cohort 1936. *Intelligence* 2011; **39**: 405–417.
- 72 Deary IJ, Der G, Ford G. Reaction times and intelligence differences: a population-based cohort study. *Intelligence* 2001; **29**: 389–399.
- 73 Der G, Deary IJ. Age and sex differences in reaction time in adulthood: results from the United Kingdom Health and Lifestyle Study. *Psychol Aging* 2006; **21**: 62–73.
- 74 Shakeshaft NG, Trzaskowski M, McMillan A, Rimpfeld K, Krapohl E, Haworth CM *et al*. Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS One* 2013; **8**: e80341.
- 75 Rimpfeld K, Kovas Y, Sale PS, Plomin R. Pleiotropy across academic subjects at the end of compulsory education. *Sci Rep* 2015; **5**: 11713.
- 76 Deary IJ, Strand S, Smith P, Fernandes C. Intelligence and educational achievement. *Intelligence* 2007; **35**: 13–21.
- 77 Krapohl E, Rimpfeld K, Shakeshaft BG, Trzaskowski M, McMillan A, Pingault JB *et al*. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc Natl Acad Sci USA* 2014; **111**: 15273–15278.
- 78 Chabris CF, Lee JJ, Benjamin DJ, Beauchamp JP, Glaeser EL, Borst G *et al*. Why it is hard to find genes associated with social science traits: theoretical and empirical considerations. *Am J Pub Health* 2013; **103**: S152–S166.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)