Haplotype-based association analysis of general cognitive ability in Generation Scotland, the English Longitudinal Study of Ageing, and UK Biobank

Howard, David M.; Adams, Mark J.; Clarke, Toni-Kim; Wigmore, Eleanor M.; Zeng, Yanni; Hagenaars, Saskia P.

Published in:
Wellcome Open Research

DOI:
10.12688/wellcomeopenres.12171.1

Publication date:
2017

RESEARCH ARTICLE

Haplotype-based association analysis of general cognitive ability in Generation Scotland, the English Longitudinal Study of Ageing, and UK Biobank [version 1; referees: 2 approved]

David M. Howard [1], Mark J. Adams[1], Toni-Kim Clarke [1], Eleanor M. Wigmore[1], Yanni Zeng[1,2], Saskia P. Hagenaars[1,3,4], Donald M. Lyall[5], Pippa A. Thomson[3,6], Kathryn L. Evans[3,6], David J. Porteous[6], Reka Nagy[2], Caroline Hayward [2,7], Chris S. Haley[2], Blair H. Smith[7,8], Alison D. Murray[7,9], G. David Batty[10], Ian J. Deary[3,4,7], Andrew M. McIntosh [1,3,7]

[1]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK
[2]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[3]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
[4]Department of Psychology, University of Edinburgh, Edinburgh, UK
[5]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
[6]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[7]Generation Scotland, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[8]Division of Population Health Sciences, University of Dundee, Dundee, UK
[9]Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK
[10]Department of Epidemiology and Public Health, University College London, London, UK

v1 **First published:** 10 Aug 2017, **2**:61 (doi: 10.12688/wellcomeopenres.12171.1)
Latest published: 10 Aug 2017, **2**:61 (doi: 10.12688/wellcomeopenres.12171.1)

Abstract

Background: Cognitive ability is a heritable trait with a polygenic architecture, for which several associated variants have been identified using genotype-based and candidate gene approaches. Haplotype-based analyses are a complementary technique that take phased genotype data into account, and potentially provide greater statistical power to detect lower frequency variants.
Methods: In the present analysis, three cohort studies (n_{total} = 48,002) were utilised: Generation Scotland: Scottish Family Health Study (GS:SFHS), the English Longitudinal Study of Ageing (ELSA), and the UK Biobank. A genome-wide haplotype-based meta-analysis of cognitive ability was performed, as well as a targeted meta-analysis of several gene coding regions.
Results: None of the assessed haplotypes provided evidence of a statistically significant association with cognitive ability in either the individual cohorts or the meta-analysis. Within the meta-analysis, the haplotype with the lowest observed *P*-value overlapped with the D-amino acid oxidase activator (*DAOA*) gene coding region. This coding region has previously been associated with bipolar disorder, schizophrenia and Alzheimer's disease, which have all been shown to impact upon cognitive ability. Another potentially interesting region highlighted within the current genome-wide association analysis (GS:SFHS: $P = 4.09 \times 10^{-7}$), was the butyrylcholinesterase (*BCHE*) gene coding region. The

Open Peer Review

Referee Status: ✓ ✓

	Invited Referees 1	2
version 1 published 10 Aug 2017	✓ report	✓ report

1 **Antony Payton**, University of Manchester, UK

2 **Heiner Rindermann**, Chemnitz University of Technology, Germany

Discuss this article

Comments (0)

protein encoded by *BCHE* has been shown to influence the progression of Alzheimer's disease and its role in cognitive ability merits further investigation. Conclusions: Although no evidence was found for any haplotypes with a statistically significant association with cognitive ability, our results did provide further evidence that the genetic variants contributing to the variance of cognitive ability are likely to be of small effect.

Corresponding author: David M. Howard (d.howard@ed.ac.uk)

Author roles: Howard DM: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Adams MJ**: Conceptualization, Data Curation, Methodology, Writing – Review & Editing; **Clarke TK**: Data Curation, Methodology, Writing – Review & Editing; **Wigmore EM**: Methodology, Writing – Review & Editing; **Zeng Y**: Data Curation, Methodology, Writing – Review & Editing; **Hagenaars SP**: Data Curation, Writing – Review & Editing; **Lyall DM**: Data Curation, Writing – Review & Editing; **Thomson PA**: Conceptualization, Methodology, Supervision, Writing – Review & Editing; **Evans KL**: Investigation, Methodology, Writing – Review & Editing; **Porteous DJ**: Supervision, Writing – Review & Editing; **Nagy R**: Investigation, Writing – Review & Editing; **Hayward C**: Data Curation, Methodology, Supervision, Writing – Review & Editing; **Haley CS**: Methodology, Supervision, Writing – Review & Editing; **Smith BH**: Resources, Writing – Review & Editing; **Murray AD**: Resources, Writing – Review & Editing; **Batty GD**: Data Curation, Writing – Review & Editing; **Deary IJ**: Data Curation, Methodology, Resources, Supervision, Writing – Review & Editing; **McIntosh AM**: Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing

Competing interests: DJP and IJD are participants in UK Biobank. The authors report that no other conflicts of interest exist.

How to cite this article: Howard DM, Adams MJ, Clarke TK *et al.* **Haplotype-based association analysis of general cognitive ability in Generation Scotland, the English Longitudinal Study of Ageing, and UK Biobank [version 1; referees: 2 approved]** Wellcome Open Research 2017, **2**:61 (doi: 10.12688/wellcomeopenres.12171.1)

Grant information: This work was supported by the Wellcome Trust [104036]; Medical Research Council and the Biotechnology and Biological Sciences Research Council [MR/K026992/1]; the Scottish Government Health Directorate [CZD/16/6]; the Scottish Funding Council [HR03006]; the Dr. Mortimer and Theresa Sackler Foundation; and the China Scholarship Council
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

First published: 10 Aug 2017, **2**:61 (doi: 10.12688/wellcomeopenres.12171.1)

Introduction

Cognitive ability facilitates the way in which we understand, interpret and interact with the world around us, and encompasses a broad range of neuropsychological skills, such as reasoning, various forms of memory, literacy, numeracy, logic, decision making, knowledge, and processing speed. There are positive correlations between each of these skills[1], and an individual's aptitude for each skill can be quantified by completing specifically designed, validated and standardised tests. The results obtained using these tests are commonly combined to form an overall general cognitive function ('*g*' or general intelligence) score. The heritability of *g* generally increases with age, with estimates ranging from 30 – 80%[2,3]. Several large, well-powered studies[4–8] have reported a number of genome-wide significant associations for cognitive phenotypes using genotype data. Despite this, genotype-based analyses using single nucleotide polymorphism (SNP) data are unlikely to be able to fully capture the variation in the regions adjacent to the typed markers. This will be especially true for untyped or rare variants, or those variants that are in weak linkage disequilibrium (LD) with the SNPs found on common genotyping arrays. Haplotypes have the additional benefit of incorporating information from multiple variants where the DNA strand has been assigned.

Haplotype-based analyses of cognitive ability have focused on a number of specific gene coding regions: brain-derived neurotrophic factor (*BDNF*)[9,10], D-amino acid oxidase activator (*DAOA*)[11,12] and apolipoprotein E (*APOE*)[13,14]. In the present analysis, these three regions will be assessed using the three available cohort studies, along with a genome-wide haplotype-based association analysis of cognitive ability. The Generation Scotland: Scottish Family Health Study (GS:SFHS) will be used as the discovery cohort, with the English Longitudinal Study of Ageing (ELSA) and UK Biobank used as replication cohort studies along with a meta-analyses of all three cohorts.

Materials and methods

Discovery cohort

Generation Scotland: Scottish Family Health Study (GS:SFHS). GS:SFHS[15,16] is a population and family-based cohort study of 23,960 individuals, of whom 20,195 were genotyped using the Illumina OmniExpress BeadChip (706,786 SNPs). Within GS:SFHS, there were 4,933 families containing at least two related individuals, including 1,799 families with two members, 1,216 families with three members and 829 families with four members, with the largest family containing 31 individuals. There were 1,789 individuals with no other family members in the cohort.

For quality control, individuals with a genotype call rate < 98% or who were identified as population outliers[17] through principal component analysis were removed, leaving 19,904 individuals. Quality control was also applied to the genomic data, with SNPs with a call rate < 98%, minor allele frequency (MAF) < 0.01 or that deviating from Hardy-Weinberg equilibrium ($P < 10^{-6}$) removed. This left a total of 561,125 autosomal SNPs.

Replication cohorts

English Longitudinal Study of Ageing (ELSA). ELSA[18] is a population-based cohort study consisting of 11,391 individuals, of which 7,597 were genotyped using the Illumina Omni 2.5–8 array (≈ 2.5M SNPs). SNPs which overlapped with the discovery sample were extracted, and individuals that reported a non-Caucasian ethnicity were removed to maximise homogeneity within the sample. This left 7,452 individuals with variant calls for 554,079 SNPs for analysis. There was no evidence of overlapping individuals between ELSA and GS:SFHS using a checksum-based approach, whereby a total of 500 randomly selected genome-wide SNPs, present across both cohort studies, were assigned to 10 equal-sized batches. A checksum was calculated using the cksum unix command for each individual and for each batch. If an individual in one cohort study had the same checksum for a specific batch as an individual in the other cohort, then this provided evidence of overlap between those two individuals (personal communication with Stephan Ripke).

UK Biobank. UK Biobank[19] is a population-based cohort study consisting of 152,249 genotyped individuals with imputed genomic data for 72,355,667 variants[20]. Individuals who reported a non-white British ethnicity or were identified as overlapping with either GS:SFHS (n = 174) or ELSA (n = 85), using the checksum-based approach described previously, were removed, leaving 119,832 individuals. Imputed variants with an infoscore ≥ 0.8, that were also genotyped in GS:SFHS, were extracted from the UK Biobank data, which identified 555,782 variants in common between the two cohorts.

Genotype phasing and haplotype formation

Phasing of the genotype data within each cohort study was conducted using SHAPEIT v2.r837[21]. Genome-wide phasing was applied to the GS:SFHS discovery cohort. Within the replication cohort studies, phasing was conducted across a 50Mb window centred on haplotypes with $P < 10^{-6}$ in the genome-wide analysis of the discovery cohort study, and the *BDNF*, *DAOA* and *APOE* gene coding regions. To improve phasing accuracy, the number of conditioning states per SNP was increased from the default of 100 states to 200 states. The default effective population size for European populations of 15,000 was used across the three cohorts. A 5Mb window size was used to conduct the phasing within GS:SFHS (rather than the default window size of 2Mb used for ELSA and UK Biobank), as this has been shown to be advantageous when larger amounts of identity by descent (IBD) sharing are present[21]. The extensive family structure within GS:SFHS also meant the duoHMM method could be applied to that cohort. The duoHMM method combined the results of a MCMC algorithm with pedigree information to improve phasing accuracy[22]. HapMap phase II b37[23] was used to calculate the recombination rates between SNPs during phasing, and for the subsequent partitioning of the phased data into haplotypes.

Window sizes of 1cM, 0.5cM and 0.25cM were used to determine the SNPs included within each haplotype[24]. A sliding window was used, sliding the window along a quarter of the respective window size. This produced a total of 97,333 windows with a mean number of SNPs per window of 157, 79 and 34 for the 1cM, 0.5cM and 0.25cM windows, respectively. The haplotype positions reported subsequently are given in base pair (bp) position (using GRCh37) and correspond to the outermost SNPs located within each haplotype. Those haplotypes containing less than

5 SNPs, or with a frequency < 0.005 or that deviating from Hardy-Weinberg equilibrium ($P < 10^{-6}$) were not assessed, but they were included as part of the alternative haplotype for the assessment of the remaining haplotypes. Following quality control there were 2,618,094 haplotypes for further analysis.

To estimate the correction required for multiple testing, the clump command within Plink v1.90[25] was used to determine the number of independently segregating haplotypes. An LD r^2 threshold of 0.4 was used to classify a haplotype as independent and at this threshold there were 1,070,216 independently segregating haplotypes in the discovery cohort study. Therefore, a Bonferroni correction required that $P < 5 \times 10^{-8}$ for genome-wide significance. This was in alignment with the conventional level for significance used for sequence and SNP-based genome-wide association studies[26]. Therefore in the present analysis, and for future genome-wide haplotype-based analyses using cohorts similar to GS:SFHS, the conventional *P*-value for significance can be applied.

General cognitive ability

Within each cohort study, a principal component analysis was used to determine a general cognitive ability score (*g*). This was calculated using the first unrotated principal component from the series of cognitive tests conducted within each cohort. The loadings used within each cohort are provided in Supplementary Table S1. The study demographics of each cohort for individuals for which *g* could be calculated are provided in Table 1. The GCTA-GREML[27] method was used to calculate SNP-based estimates for the heritability of *g*.

Generation Scotland: Scottish Family Health Study (GS:SFHS). The following tests were used within GS:SFHS to calculate *g*: logical memory, verbal fluency, digit symbol-coding, and vocabulary. Logical memory was assessed using the Wechsler Memory Scale III[28]. Verbal fluency was measured using a phonemic fluency test, requiring the participant to name as many words as possible beginning with a particular letter (C, F, and L were used) within a given timeframe[29]. Digit symbol-coding was assessed using the Wechsler Adult Intelligence Scale III[29]. Vocabulary was assessed using the Mill Hill Vocabulary Scale senior and junior synonyms combined[30]. Additional information regarding the cognitive ability variables available within GS:SFHS has been published previously[14,15,31]. *g* explained 0.43 of the variance across the four tests and was available for 19,326 individuals.

Table 1. Study demographics of Generation Scotland: Scottish Family Health Study (GS:SFHS), English Longitudinal Study of Ageing (ELSA) and UK Biobank for individuals with a general intelligence score.

	GS:SFHS	ELSA	UK Biobank
N	19,326	5,876	22,800
Males/Females	7,929/11,397	2,679/3,197	10,665/12,135
Age Range	18 – 94	31 – 90	40 – 75
Mean Age (s.dev.)	47.2 (14.9)	63.3 (9.4)	56.4 (7.7)

English Longitudinal Study of Ageing (ELSA). The first wave of the cognitive tests conducted by ELSA were used to calculate *g* for this cohort: processing speed, verbal memory and verbal fluency. Processing speed was calculated using a letter cancellation task with participants searching a large grid of letters for the letters P and W and crossing those out. Verbal memory was assessed using a ten-word list-learning task. Verbal fluency was measured by the number of different animal species that could be named in one minute. Further information regarding these cognitive tests is provided elsewhere[32,33]. There were 5,876 individuals for which *g* could be calculated, with *g* explaining 0.49 of the total variance across the three cognitive tests.

UK Biobank. The touchscreen cognitive tests conducted as part of the online follow-up within UK Biobank were used to derive *g*. Some of these tests have yet to be reported elsewhere and are therefore covered in greater detail here. The following tests were used within this cohort study: fluid intelligence test (UK Biobank Field 20191), trail making test (mean of UK Biobank Fields 20156 and 20157), symbol digit substitution test (UK Biobank Field 20159) and numeric memory test (UK Biobank Field 20240). The fluid intelligence test consisted of 13 multiple-choice questions to be answered within two minutes, with a score based on the number of correct answers. For the trail making test participants were firstly presented with a screen containing a series of numbers from 1 to 25, each contained within a circle. Starting with the circle containing the number 1, the participants then had to use the computer mouse to click on the numbers in ascending order. Secondly, the participants were presented with circles containing the numbers 1 to 13 and the letters A to L. For this test the participants had to click the circles in the order 1, A, 2, B, 3, C, 4, D, etc. For both the trail making tests the time taken to complete each test was recorded, with the log of the mean time across the two tests taken as the final score for this test. The symbol digit coding test consisted of a series of eight symbols that corresponded to eight numbers. The participants were then repetitively presented with eight symbols in a specific order that required recoding to their numerical equivalents. The number of correctly recoded sequences within one minute was recorded. The numeric memory test began with a two-digit number being presented, after a short delay the participant was then required to enter the number presented. The length of the number presented was then incremented by one digit each time with the participant required to recall the full number correctly, up to a maximum of 12 digits. The maximum number of digits recalled successfully was recorded. The proportion of variance explained by *g* across the four tests was 0.51 and was available for 22,800 individuals. The proportion of variance explained by *g* within the online follow-up was greater than that reported (≈ 0.4) by Lyall, Cullen[34] for the original cognitive tests conducted within UK Biobank.

Statistical analysis

Discovery cohort. A genome-wide haplotype-based association analysis was conducted within GS:SFHS using a mixed linear model within GCTA v1.25.0[35]:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{v} + \boldsymbol{\varepsilon}$$

where **y** was the vector of observations for *g*. **β** was the matrix of fixed effects, including haplotype, sex and age. A SNP-based genomic relationship matrix[27] (**G**) using the 'leave one chromosome out' methodology[35], which excluded the chromosome of the assessed haplotype, was fitted as a random effect, **u**, taking into account the genomic relationships as MVN $(0, \mathbf{G}\sigma^2_{\mathbf{u}})$. **v** was a random effect fitting a second genomic relationship matrix $\mathbf{G}_t$ as MVN $(0, \mathbf{G}_t\sigma^2_{\mathbf{v}})$, which modelled only the more closely related individuals[36]. $\mathbf{G}_t$ was identical to **G**, except that off-diagonal elements < 0.05 were set to 0. **X**, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ were the corresponding incidence matrices. **ε** was the vector of residual effects and was assumed to be normally distributed as MVN $(0, \mathbf{I}\sigma^2_{\varepsilon})$.

GS:SFHS is a family-based cohort and therefore LD score regression[37] was used to test for the existence of population stratification by examining the summary statistics obtained from the above mixed model. The fitting of a single genomic relationship matrix, **G**, provided evidence of significant population stratification (intercept = 1.051 ± 0.004). Whilst the simultaneous fitting of the matrices **G** and to $\mathbf{G}_t$ together produced no evidence of population stratification (intercept = 0.998 ± 0.003), hence the fitting of two matrices for GS:SFHS.

Replication cohorts. A mixed linear model was used to assess the haplotypes in ELSA and UK Biobank which were identified in the GS:SFHS discovery cohort study with $P < 10^{-6}$ and those haplotypes in GS:SFHS that overlapped with the *BDNF*, *DAOA* and *APOE* gene coding regions. This was conducted using GCTA v1.25.0[35]:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{u} + \boldsymbol{\varepsilon}$$

where **y** was the vector of binary observations for g. **β** was the matrix of fixed effects, including haplotype, sex and age, and for UK Biobank, genotyping batch and recruitment centre were also fitted. **u** was fitted as a random effect taking into account the SNP-based genomic relationships as MVN $(0, \mathbf{G}\sigma^2_{\mathbf{u}})$ and also implemented the 'leave one chromosome out' methodology[35]. **X** and $\mathbf{Z}_1$ were the corresponding incidence matrices and **ε** was the vector of residual effects and was assumed to be normally distributed as MVN $(0, \mathbf{I}\sigma^2_{\varepsilon})$. Replication success was judged on the statistical significance of each haplotype using an inverse variance-weighted meta-analysis across all three cohorts conducted with Metal[38].

Results

A genome-wide haplotype-based association analysis for general cognitive ability, using a principal component derived measure of *g*, was conducted using 2,618,094 haplotypes within the GS:SFHS discovery cohort study. A genome-wide Manhattan plot of $-\log_{10}$ *P*-values is provided in Figure 1, with a q-q plot provided in Supplementary Figure S1. No haplotypes exceeded the genome-wide significance threshold ($P < 5 \times 10^{-8}$) for an association with *g*. Within the discovery cohort study, 12 haplotypes had

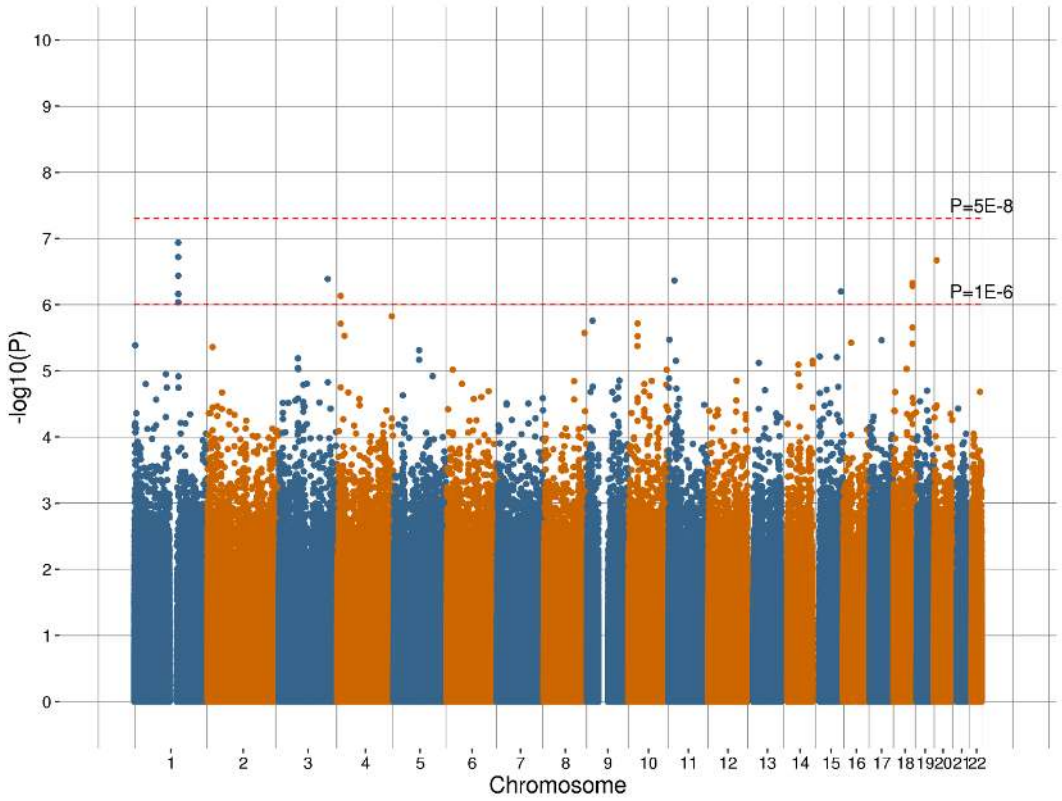

Figure 1. Manhattan plot representing the $-\log_{10}$ *P*-values for an association between each assessed haplotype and cognitive score in the Generation Scotland: Scottish Family Health Study cohort study.

$P < 10^{-6}$, and replication was sought for these 12 haplotypes within ELSA and UK Biobank. Summary statistics regarding each cohort study and the meta-analysis of these haplotypes (after applying an LD r^2 threshold of 0.4 to identify those that are independently segregating) are provided in Table 2. The frequencies of the haplotypes within each cohort, for the seven independently segregating haplotypes with $P < 10^{-6}$ in the discovery cohort, along with the protein coding genes that these haplotypes overlapped, are provided in Supplementary Table S2.

Of the 12 haplotypes with $P < 10^{-6}$ in GS:SFHS, none were nominally significant ($P \geq 0.05$) in ELSA. Within UK Biobank the only haplotype to be nominally significantly ($P < 0.05$) associated with *g* was located on chromosome 11 and this was in the opposite direction to that observed for GS:SFHS. The smallest *P*-value (1.46×10^{-3}) observed within the genome-wide meta-analysis was located on chromosome 18 and although neither of the replication cohort studies were nominally significant, their effects were in the same direction as that observed within GS:SFHS. The genetic variance explained by each of the haplotypes within GS:SFHS was small, ranging from $3.93 \times 10^{-3} - 4.63 \times 10^{-3}$. A power analysis revealed that the sample sizes for the replication cohorts were large enough to provide statistical power in excess of 0.99, assuming an effect size equivalent to that observed in the discovery cohort study.

The SNP-based heritability of *g* was calculated using GCTA-GREML[27] and was 0.41 (s.e = 0.05) for GS:SFHS, 0.17 (s.e. = 0.06) for ELSA, and 0.21 (s.e. = 0.02) for UK Biobank. The heritability of *g* within GS:SFHS was calculated using an unrelated subsample of 7 388 individuals, whereby one of a pair of individuals was removed if they shared a genotype-based relatedness of > 0.025.

BDNF, DAOA and *APOE* gene coding regions

None of the haplotypes overlapping the *BDNF, DAOA* and *APOE* gene coding regions were statistically significant at the genome-wide level ($P \geq 5 \times 10^{-8}$) in the meta-analysis or in the single cohort analyses. The top five independently segregating haplotypes (following the application of an LD r^2 threshold of 0.4) in terms of statistical significance achieved in the meta-analysis for each of the gene coding regions are shown in Table 3. There were 214 haplotypes that overlapped the *BDNF* gene coding region and the lowest *P*-value obtained in the meta-analysis was 1.35×10^{-3} for a haplotype with a positive effect (beta = 0.31 ± 0.10) on *g*. The *DAOA* gene coding region overlapped with 410 assessed haplotypes, with the lowest *P*-value = 1.53×10^{-5} within the meta-analysis for a haplotype with a positive effect (beta = 0.20 ± 0.05) on *g*. Overlapping the *APOE* gene coding region there were 325 assessed haplotypes, of which the lowest observed *P*-value in the meta-analysis was 7.50×10^{-4} for a haplotype with a positive effect (beta = 0.18 ± 0.05).

Discussion

Twelve haplotypes were identified in the GS:SFHS discovery cohort study with a *P*-value $< 10^{-6}$ for an association with *g*, although none of these reached genome-wide significance ($P > 5 \times 10^{-8}$). Replication of these twelve haplotypes was sought and not found within the ELSA and UK Biobank cohort studies. Both of these cohorts were sufficiently powered cohorts to detect effects of the sizes observed within GS:SFHS, assuming that the haplotypes were in linkage equilibrium with the causal variant. Therefore, despite SNP-based heritability estimates ranging from 0.17 to 0.41 for *g* across the three cohort studies, there was no evidence for any haplotypes significantly associated with cognitive ability.

The haplotypes with $P < 10^{-6}$ within the discovery cohort study overlapped with a number of gene coding regions. In terms of biological viability the most notable of these haplotypes was located on chromosome 3 that overlapped with the coding region for the butyrylcholinesterase (*BCHE*) gene. *BCHE* has been shown to have a role in cognitive ability within humans[39,40] as well as rodents[41,42]. SNP variants close to this coding region, which overlapped with the haplotype on chromosome 3, have also been shown to be significantly associated ($P = 2.69 \times 10^{-8}$) with the cortical deposition of amyloid-β peptide[43]. This deposition is thought to

Table 2. Independently segregating (linkage disequilibrium r^2 threshold of 0.4) haplotypes sorted by *P*-value obtained in the meta-analysis and with a *P*-value < 10^{-6} for an association with cognitive ability within the discovery cohort study, Generation Scotland: Scottish Family Health Study (GS:SFHS).

Haplotype		GS:SFHS		ELSA		UK Biobank		Meta-analysis	
Chr	Position (bp)	Beta (s.e.)	*P*-value	Beta (s.e.)	*P*-value	Beta (s.e.)	*P*-value	Direction	*P*-value
18	64252341 - 64568113	0.23 (0.05)	5.21×10^{-7}	0.08 (0.06)	0.17	0.01 (0.02)	0.54	+++	0.001
3	165337109 - 166522847	0.60 (0.12)	4.09×10^{-7}	-0.02 (0.15)	0.9	0.06 (0.06)	0.36	+-+	0.003
20	9288522 - 9726640	0.55 (0.11)	2.13×10^{-7}	-0.003 (0.06)	0.96	0.04 (0.05)	0.38	+-+	0.008
1	150165849 - 151140732	0.51 (0.10)	9.20×10^{-7}	0.12 (0.14)	0.37	0.002 (0.06)	0.97	+++	0.01
4	11448182 - 11547967	0.32 (0.06)	7.36×10^{-7}	0.11 (0.07)	0.96	0.01 (0.03)	0.83	+++	0.04
11	20184958 - 20297638	-0.56 (0.11)	4.31×10^{-7}	-0.03 (0.13)	0.84	0.12 (0.06)	0.04	--+	0.6
15	94701431 - 94729657	-0.27 (0.05)	6.33×10^{-7}	-0.10 (0.07)	0.14	0.02 (0.01)	0.16	--+	0.85

Beta values, standard errors and *P*-values are given for GS:SFHS, English Longitudinal Study of Ageing (ELSA), UK Biobank and a meta-analysis of all three cohort studies. Genomic location is determined by position on the GRCh37 assembly.

Table 3. Independently segregating (linkage disequilibrium r^2 threshold of 0.4) haplotypes overlapping the brain-derived neurotrophic factor (BDNF), D-amino acid oxidase activator (DAOA) and apolipoprotein E (APOE) gene coding regions.

Haplotype		GS:SFHS		ELSA		UK Biobank		Meta-analysis	
Gene	**Chr:Position (bp)**	**Beta (s.e.)**	***P*-value**	**Beta (s.e.)**	***P*-value**	**Beta (s.e.)**	***P*-value**	**Direction**	***P*-value**
BDNF	11:27337843-27778592	0.34 (0.12)	0.007	0.27 (0.15)	0.08	na	na	++?	0.001
	11:27444517-27787783	0.31 (0.12)	0.01	0.24 (0.15)	0.11	na	na	++?	0.003
	11:27337843-27778592	0.22 (0.09)	0.01	0.10 (0.09)	0.27	na	na	++?	0.009
	11:27662826-27990119	0.25 (0.09)	0.006	0.07 (0.10)	0.47	na	na	++?	0.01
	11:27020461-27749725	-0.28 (0.11)	0.01	-0.10 (0.10)	0.31	na	na	--?	0.02
DAOA	13:106140780-106393146	0.25 (0.11)	0.03	0.16 (0.14)	0.23	0.20 (0.06)	3.54×10^{-4}	+++	1.53×10^{-5}
	13:106098389-106240125	0.28 (0.11)	0.009	0.08 (0.11)	0.47	0.13 (0.05)	0.005	+++	2.63×10^{-4}
	13:106140780-106240125	0.22 (0.10)	0.02	0.08 (0.10)	0.42	0.12 (0.04)	0.005	+++	4.03×10^{-4}
	13:106066286-106154577	-0.22 (0.12)	0.07	-0.06 (0.14)	0.67	-0.22 (0.07)	0.002	---	5.91×10^{-4}
	13:106065361-106133365	-0.18 (0.11)	0.12	-0.04 (0.13)	0.75	-0.20 (0.06)	0.001	---	6.50×10^{-4}
APOE	19:45290685-45422561	0.28 (0.11)	0.009	0.18 (0.13)	0.16	0.14 (0.07)	0.05	+++	7.50×10^{-4}
	19:45318153-45422561	0.27 (0.09)	0.003	0.20 (0.10)	0.05	0.06 (0.05)	0.28	+++	0.002
	19:45389224-45548502	0.14 (0.08)	0.07	0.11 (0.09)	0.21	0.08 (0.04)	0.04	+++	0.004
	19:45390685-45422561	0.09 (0.13)	0.45	-0.15 (0.13)	0.26	-0.17 (0.06)	0.004	+--	0.01
	19:45351746-45422561	0.39 (0.10)	1.24×10^{-4}	-0.09 (0.11)	0.4	0.08 (0.05)	0.14	+-+	0.01

Beta values, standard errors and *P*-values are given for Generation Scotland: Scottish Family Health Study (GS:SFHS), English Longitudinal Study of Ageing (ELSA), UK Biobank and a meta-analysis of all three cohort studies. There were no UK Biobank individuals that carried the shown BDNF overlapping haplotypes. Haplotypes are sorted by *P*-value obtained in the meta-analysis within each gene coding region. Genomic location is determined by position on the GRCh37 assembly.

be an initiating factor in the pathology of Alzheimer's disease[44,45], which has a known impact on cognitive ability. Furthermore, the *BCHE*-K variant (rs1803274) has been shown to have an effect on the progression of Alzheimer's disease[46,47] and an interaction with the *APOE* ε4 allele among those with late-onset of the disease[48]. The *BCHE*-K variant was not genotyped within GS:SFHS but it is located within the bounds of the haplotype on chromosome 3. This haplotype was analysed and not found to be associated with Alzheimer's disease ($P \geq 0.05$) within GS:SFHS, using the same mixed linear model described previously and self-declared Alzheimer's disease as the phenotype. However, the prevalence of the disease in this cohort (0.14%) is likely to have limited the power to detect an effect.

The targeted meta-analyses of the *BDNF*, *DAOA* and *APOE* gene coding regions did not provide evidence of genome-wide significant haplotypes ($P \geq 5 \times 10^{-8}$) associated with cognitive ability. The *BDNF* region yielded several haplotypes which were more statistically significant than those found by Wilkosc, Szalkowska[9] or Warburton, Miyajima[10]. *BDNF* is involved in the development of synaptic connectivity in the central nervous system[49] and therefore represents a potential source of cognitive score variance. The most significant haplotype ($P = 1.53 \times 10^{-5}$) identified across all meta-analyses was in the *DAOA* coding region. SNP variants located within the *DAOA* gene have also been associated with diseases related to the brain: bipolar disorder[50], Alzheimer's disease[51] and, potentially, schizophrenia[52]. These diseases are known to be associated with decrements in cognitive ability. Haplotypes within the *APOE* gene coding region have been studied previously within GS:SFHS[14], although the haplotypes examined previously were considerably shorter, formed of two variants and used the cognitive tests individually rather than forming an overall *g* score. The *P*-value of the most significant haplotype in the *APOE* region in the present analysis was stronger than the haplotypes assessed by Marioni, Campbell[14], but was not genome-wide significant ($P \geq 5 \times 10^{-8}$).

The cohort studies selected for analysis should be relatively homogenous, as they are a subset of the British population, this can be observed by the consistency of the haplotype frequencies shown in Supplementary Table S2. However, there were some differences in the cognitive tests applied between the studies. The size of the present analysis is comparable number to that of the genotyped-based genome-wide association study of cognitive ability conducted by the CHARGE consortium[4]. Their paper drew the conclusion that there were likely to be many genes of small effect contributing to the genetic variance underlying cognitive ability. Based on the observed heritability of the trait, but a lack of genome-wide significant haplotypes in the present analyses, this conclusion continues to hold true.

Conclusions

None of the haplotypes analysed in this study achieved genome-wide significance ($P \geq 5 \times 10^{-8}$) for an association with cognitive

ability within any of the cohort studies, or in the meta-analysis. The genome-wide analysis identified a haplotype within the *BCHE* gene coding region which may play a role in cognitive ability and this warrants further analysis. Although haplotypes should allow the detection of signals from rarer causal variants compared to a typical genotype-based analysis, there was no evidence for genome-wide significant haplotypes for the window sizes tested. Potentially shorter and therefore more common haplotypes could be assessed, however to detect rarer genetic contributions to highly polygenic traits such as cognitive ability, there remains a requirement for larger sample sizes.

Data availability

Due to the confidential nature of the genetic data and cognitive test scores of participants, it is not possible to publically share the data on which our analysis was based. Generation Scotland (GS) data is available on request to: access@generationscotland.org, with further information available from http://www.ed.ac.uk/generation-scotland. Each application requires the completion of a data and materials transfer agreement, the conditions of which be determined on a case by case basis. GS has Research Tissue Bank status, and the GS Access Committee reviews applications to ensure that they comply with legal requirements, ethics and participant consent. UK Biobank data is available for health related research on request to: access@ukbiobank.ac.uk, with further information relating to data access available from http://www.ukbiobank.ac.uk/register-apply. The English Longitudinal Study of Ageing data is available on request to: n.rogers@ucl.ac.uk, with further information regarding data access available from https://www.elsa-project.ac.uk.

Ethical statement

Ethics approval for the Generation Scotland study was given by the NHS Tayside committee on research ethics (reference 05/S1401/8). The UK Biobank study was conducted under generic approval from the NHS National Research Ethics Service (approval letter dated 17th June 2011, Ref 11/NW/0382). Ethical approval for the English Longitudinal Study of Ageing was obtained from the London Multi-Centre Research Ethics Committee. All participants gave full informed written consent to participate within each study.

Competing interests

DJP and IJD are participants in UK Biobank. The authors report that no other conflicts of interest exist.

Grant information

This work was supported by the Wellcome Trust [104036]; Medical Research Council and the Biotechnology and Biological Sciences Research Council [MR/K026992/1]; the Scottish Government Health Directorate [CZD/16/6]; the Scottish Funding Council [HR03006]; the Dr. Mortimer and Theresa Sackler Foundation; and the China Scholarship Council.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. This research has been conducted using the UK Biobank Resource – application number 4844. We are grateful to both the UK Biobank and the English Longitudinal Study of Ageing and their participants.

Supplementary material

Supplementary Table S1. Loadings used for each of the cognitive tests to calculate the general intelligence (*g*) score within Generation Scotland: Scottish Family Health Study (GS:SFHS), the English Longitudinal Study of Ageing (ELSA) and UK Biobank.

Click here to access the data.

Supplementary Table S2. Haplotype frequency and overlap with gene coding regions of the independently segregating (linkage disequilibrium r^2 threshold of 0.4) haplotypes with a *P*-value < 10^{-6} for an association with cognitive ability within the discovery cohort study, Generation Scotland: Scottish Family Health Study (GS:SFHS). Haplotype frequencies are also provided for the English Longitudinal Study of Ageing (ELSA) and UK Biobank cohort studies. Haplotypes are sorted by chromosome with genomic locations determined by position on the GRCh37 assembly.

Click here to access the data.

Supplementary Figure S1. Q-Q plot representing the observed $-\log_{10}$ *P*-values against the expected $-\log_{10}$ *P*-values for an association between each assessed haplotype and cognitive score in the Generation Scotland: Scottish Family Health Study cohort study.

Click here to access the data.

References

1. Johnson W, Nijenhuis JT, Bouchard TJ Jr: **Still just 1 *g*: consistent results from five test batteries.** *Intelligence.* 2008; **36**(1): 81–95.
Publisher Full Text
2. Deary IJ, Johnson W, Houlihan LM: **Genetic foundations of human intelligence.** *Hum Genet.* 2009; **126**(1): 215–232.
PubMed Abstract | Publisher Full Text
3. Plomin R, Deary IJ: **Genetics and intelligence differences: five special findings.** *Mol Psychiatry.* 2015; **20**(1): 98–108.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Davies G, Armstrong N, Bis JC, *et al.*: **Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (*N*=53949).** *Mol Psychiatry.* 2015; **20**(2): 183–192.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Davies G, Marioni RE, Liewald DC, *et al.*: **Genome-wide association study of cognitive functions and educational attainment in UK Biobank (*N*=112 151).** *Mol Psychiatry.* 2016; **21**(6): 758–767.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Rietveld CA, Medland SE, Derringer J, *et al.*: **GWAS of 126,559 individuals identifies genetic variants associated with educational attainment.** *Science.* 2013; **340**(6139): 1467–1471.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Ibrahim-Verbaas CA, Bressler J, Debette S, *et al.*: **GWAS for executive function and processing speed suggests involvement of the *CADM2* gene.** *Mol Psychiatry.* 2016; **21**(2): 189–197.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Trampush JW, Yang ML, Yu J, *et al.*: **GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium.** *Mol Psychiatry.* 2017; **22**(3): 336–345.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Wiłkość M, Szałkowska A, Skibińska M, *et al.*: ***BDNF* gene polymorphisms and haplotypes in relation to cognitive performance in Polish healthy subjects.** *Acta Neurobiol Exp (Wars).* 2016; **76**(1): 43–52.
PubMed Abstract
10. Warburton A, Miyajima F, Shazadi K, *et al.*: **NRSF and BDNF polymorphisms as biomarkers of cognitive dysfunction in adults with newly diagnosed epilepsy.** *Epilepsy Behav.* 2016; **54**: 117–127.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Jansen A, Krach S, Krug A, *et al.*: **Effect of the G72 (DAOA) putative risk haplotype on cognitive functions in healthy subjects.** *BMC Psychiatry.* 2009; **9**(1): 60.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Opgen-Rhein C, Lencz T, Burdick KE, *et al.*: **Genetic variation in the DAOA gene complex: Impact on susceptibility for schizophrenia and on cognitive performance.** *Schizophr Res.* 2008; **103**(1–3): 169–177.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Prada D, Colicino E, Power MC, *et al.*: **Influence of multiple *APOE* genetic variants on cognitive function in a cohort of older men - results from the Normative Aging Study.** *BMC Psychiatry.* 2014; **14**(1): 223.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Marioni RE, Campbell A, Scotland G, *et al.*: **Differential effects of the *APOE* e4 allele on different domains of cognitive ability across the life-course.** *Eur J Hum Genet.* 2016; **24**(6): 919–923.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Smith BH, Campbell A, Linksted P, *et al.*: **Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness.** *Int J Epidemiol.* 2013; **42**(3): 689–700.
PubMed Abstract | Publisher Full Text
16. Nagy R, Boutin TS, Marten J, *et al.*: **Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants.** *Genome Med.* 2017; **9**(1): 23, in press.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Amador C, Huffman J, Trochet H, *et al.*: **Recent genomic heritage in Scotland.** *BMC Genomics.* 2015; **16**(1): 437.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Steptoe A, Breeze E, Banks J, *et al.*: **Cohort profile: the English longitudinal study of ageing.** *Int J Epidemiol.* 2013; **42**(6): 1640–1648.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Allen NE, Sudlow C, Peakman T, *et al.*: **UK biobank data: come and get it.** *Sci Transl Med.* 2014; **6**(224): 224ed4.
PubMed Abstract | Publisher Full Text
20. Marchini J: **UK Biobank phasing and imputation documentation.** Version 1.2. 2015.
Reference Source
21. Delaneau O, Zagury JF, Marchini J: **Improved whole-chromosome phasing for disease and population genetic studies.** *Nat Methods.* 2013; **10**(1): 5–6.
PubMed Abstract | Publisher Full Text
22. O'Connell J, Gurdasani D, Delaneau O, *et al.*: **A general approach for haplotype phasing across the full spectrum of relatedness.** *PLoS Genet.* 2014; **10**(4): e1004234.
PubMed Abstract | Publisher Full Text | Free Full Text
23. International HapMap Consortium, Frazer KA, Ballinger DG, *et al.*: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature.* 2007; **449**(7164): 851–861.
PubMed Abstract | Publisher Full Text | Free Full Text
24. Browning BL, Browning SR: **Improving the accuracy and efficiency of identity-by-descent detection in population data.** *Genetics.* 2013; **194**(2): 459–471.
PubMed Abstract | Publisher Full Text | Free Full Text
25. Purcell S, Neale B, Todd-Brown K, *et al.*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet.* 2007; **81**(3): 559–575.
PubMed Abstract | Publisher Full Text | Free Full Text
26. Sham PC, Purcell SM: **Statistical power and significance testing in large-scale genetic studies.** *Nat Rev Genet.* 2014; **15**(5): 335–346.
PubMed Abstract | Publisher Full Text
27. Yang J, Benyamin B, McEvoy BP, *et al.*: **Common SNPs explain a large proportion of the heritability for human height.** *Nat Genet.* 2010; **42**(7): 565–569.
PubMed Abstract | Publisher Full Text | Free Full Text
28. Wechsler D: **Wechsler Memory Scale–Third Edition.** 1997.
Reference Source
29. Wechsler D: **Wechsler Adult Intelligence Scale–Third Edition (WAIS–III).** 1997.
Reference Source
30. Raven JC: **Mill Hill Vocabulary Scales.** *Pearson.* 2003.
Reference Source
31. Hagenaars SP, Harris SE, Clarke TK, *et al.*: **Polygenic risk for coronary artery disease is associated with cognitive ability in older adults.** *Int J Epidemiol.* 2016; **45**(2): 433–440, pii: dyv354.
PubMed Abstract | Publisher Full Text | Free Full Text
32. Llewellyn DJ, Lang IA, Langa KM, *et al.*: **Cognitive function and psychological well-being: findings from a population-based cohort.** *Age Ageing.* 2008; **37**(6): 685–689.
PubMed Abstract | Publisher Full Text | Free Full Text
33. Steel N, Huppert FA, McWilliams B, *et al.*: **Physical and cognitive function.** In *Health, Wealth and Lifestyles of the Older Population in England: the 2002 English Longitudinal Study of Ageing.* M. Marmot, *et al.*, Editors. London: Institute of Fiscal Studies. 2003; 249–300.
Reference Source
34. Lyall DM, Cullen B, Allerhand M, *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.
PubMed Abstract | Publisher Full Text | Free Full Text
35. Yang J, Zaitlen NA, Goddard ME, *et al.*: **Advantages and pitfalls in the application of mixed-model association methods.** *Nat Genet.* 2014; **46**(2): 100–106.
PubMed Abstract | Publisher Full Text | Free Full Text
36. Zaitlen N, Kraft P, Patterson N, *et al.*: **Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits.** *PLoS Genet.* 2013; **9**(5): e1003520.
PubMed Abstract | Publisher Full Text | Free Full Text
37. Bulik-Sullivan BK, Loh PR, Finucane HK, *et al.*: **LD score regression distinguishes confounding from polygenicity in genome-wide association studies.** *Nat Genet.* 2015; **47**(3): 291–295.
PubMed Abstract | Publisher Full Text | Free Full Text
38. Willer CJ, Li Y, Abecasis GR: **METAL: fast and efficient meta-analysis of genomewide association scans.** *Bioinformatics.* 2010; **26**(17): 2190–2191.
PubMed Abstract | Publisher Full Text | Free Full Text
39. Manoharan I, Kuznetsova A, Fisk JD, *et al.*: **Comparison of cognitive functions between people with silent and wild-type butyrylcholinesterase.** *J Neural Transm (Vienna).* 2007; **114**(7): 939–45.
PubMed Abstract | Publisher Full Text
40. Fiocco AJ, Nair NP, Schwartz G, *et al.*: **Influence of genetic polymorphisms in the apolipoprotein (APOE) and the butyrylcholinesterase (BCHE) gene on stress markers in older adults: A 3-year study.** *Neurobiol Aging.* 2009; **30**(6): 1001–1005.
PubMed Abstract | Publisher Full Text
41. Maurice T, Strehaiano M, Siméon N, *et al.*: **Learning performances and vulnerability to amyloid toxicity in the butyrylcholinesterase knockout mouse.** *Behav Brain Res.* 2016; **296**: 351–360.
PubMed Abstract | Publisher Full Text
42. Greig NH, Utsuki T, Ingram DK, *et al.*: **Selective butyrylcholinesterase inhibition elevates brain acetylcholine, augments learning and lowers Alzheimer beta-amyloid peptide in rodent.** *Proc Natl Acad Sci U S A.* 2005; **102**(47): 17213–17218.
PubMed Abstract | Publisher Full Text | Free Full Text
43. Ramanan VK, Risacher SL, Nho K, *et al.*: ***APOE* and *BCHE* as modulators of**

cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol Psychiatry.* 2014; **19**(3): 351–357.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Karran E, Mercken M, De Strooper B: **The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics.** *Nat Rev Drug Discov.* 2011; **10**(9): 698–712.
PubMed Abstract | Publisher Full Text

45. Selkoe DJ, Hardy J: **The amyloid hypothesis of Alzheimer's disease at 25 years.** *EMBO Mol Med.* 2016; **8**(6): 595–608.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Lane R, Feldman HH, Meyer J, *et al.*: **Synergistic effect of apolipoprotein E epsilon4 and butyrylcholinesterase K-variant on progression from mild cognitive impairment to Alzheimer's disease.** *Pharmacogenet Genomics.* 2008; **18**(4): 289–298.
PubMed Abstract | Publisher Full Text

47. Holmes C, Ballard C, Lehmann D, *et al.*: **Rate of progression of cognitive decline in Alzheimer's disease: effect of butyrylcholinesterase K gene variation.** *J Neurol Neurosurg Psychiatry.* 2005; **76**(5): 640–643.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Lehmann DJ, Johnston C, Smith AD: **Synergy between the genes for butyrylcholinesterase K variant and apolipoprotein E4 in late-onset confirmed Alzheimer's disease.** *Hum Mol Genet.* 1997; **6**(11): 1933–1936.
PubMed Abstract | Publisher Full Text

49. Cohen-Cory S, Kidane AH, Shirkey NJ, *et al.*: **Brain-derived neurotrophic factor and the development of structural neuronal connectivity.** *Dev Neurobiol.* 2010; **70**(5): 271–288.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Hukic DS, Frisén L, Backlund L, *et al.*: **Cognitive manic symptoms in bipolar disorder associated with polymorphisms in the *DAOA* and *COMT* genes.** *PLoS One.* 2013; **8**(7): e67450.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Vélez JI, Rivera D, Mastronardi CA, *et al.*: **A Mutation in *DAOA* Modifies the Age of Onset in *PSEN1* E280A Alzheimer's Disease.** *Neural Plast.* 2016; **2016**: 9760314.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Acar C, Kartalci S: **Schizophrenia: what's DAO and DAOA got to do with it?** *MOJ Proteomics Bioinformatics.* 2014; **1**(3): 00019.
Publisher Full Text

Open Peer Review

Current Referee Status: ✔ ✔

Version 1

Referee Report 15 September 2017

doi:10.21956/wellcomeopenres.13175.r26013

✔ **Heiner Rindermann**
Department of Psychology, Chemnitz University of Technology, Chemnitz, Germany

The paper tries to find genetic associations with intelligence. For this purpose, haplotypes and intelligence g factors are correlated at the individual level using three large British data sets. The study could not find significant (statistically significant, theoretically important) correlations between genomes and intelligence.

Study design, analysis etc. are well done.

I have three suggestions:

First, present your results also using easy to understand effect sizes as:
1. average correlations between certain genes/haplotypes and intelligence and
2. IQ effects in the IQ scale. If there are zero-correlations present them. If there are zero-IQ-effects present them.

Second, describe how evolutionarily diverse your samples are – are there only British people or also people from Northern Africa, Middle East, Central Asia and Sub-Saharan Africa?

Third, there are theoretically important and empirically found associations between genes (being coding or not) and intelligence at the individual and group level, e.g.: Davis *et al*, (2015), Piffer (2013), Piffer (2015) and Rindermann *et al.*(2012).

In these studies were mentioned several genes, genetic markers and haplogroups, e.g. DUF1220, COMT Val158Met and the haplogroups I, R1a, R1b, N, J1, E, T[+L], being associated to intelligence.
Do you have data on this? Can you replicate the findings?

References

1. Davis JM, Searles VB, Anderson N, Keeney J, Raznahan A, Horwood LJ, Fergusson DM, Kennedy MA, Giedd J, Sikela JM: DUF1220 copy number is linearly associated with increased cognitive function as measured by total IQ and mathematical aptitude scores.*Hum Genet*. 2015; **134** (1): 67-75 PubMed Abstract | Publisher Full Text
2. Piffer D.: Correlation of the COMT Val158Met polymorphism with latitude and a hunter-gather lifestyle suggests culture–gene coevolution and selective pressure on cognition genes due to climate.

Anthropological Science. 2013; **121** (3): 161-171 Publisher Full Text
3. Piffer D.: A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence*. 2015; **53**: 43-50 Publisher Full Text
4. Rindermann H, Woodley M.A., Stratford J.: Haplogroups as evolutionary markers of cognitive ability. *Intelligence*. 2012; **40** (4): 362-375 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
No

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 08 September 2017

doi:10.21956/wellcomeopenres.13175.r25393

Antony Payton
Centre for Epidemiology, Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK

Abstract
“Cognitive ability is a heritable trait with a polygenic architecture, for which several associated variants have been identified using genotype-based and candidate gene approaches”
A recent GWAS by Sniekers *et al* identified 336 variants in 22 genes associated with human intelligence. This is more than “several”.
Genotyping is used in candidate gene approaches. Does the author mean GWAS and candidate gene approaches?

“Haplotype-based analyses are a complementary technique that take phased genotype data into account, and potentially provide greater statistical power to detect lower frequency variants.”

Grammar needs checking.

Introduction
“Haplotypes have the additional benefit of incorporating information from multiple variants where the DNA strand has been assigned.”
What does “where the DNA strand has been assigned” mean? This is too vague.

References
1. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, Taskesen E, Hammerschlag AR, Okbay A, Zabaneh D, Amin N, Breen G, Cesarini D, Chabris CF, Iacono WG, Ikram MA, Johannesson M, Koellinger P, Lee JJ, Magnusson PKE, McGue M, Miller MB, Ollier WER, Payton A, Pendleton N, Plomin R, Rietveld CA, Tiemeier H, van Duijn CM, Posthuma D: Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence.*Nat Genet*. 2017; **49** (7): 1107-1112 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.