Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function

Gail Davies¹, Max Lam et al.#

General cognitive function is a prominent and relatively stable human trait that is associated with many important life outcomes. We combine cognitive and genetic data from the CHARGE and COGENT consortia, and UK Biobank (total N = 300,486; age 16–102) and find 148 genome-wide significant independent loci (P < 5 × 10⁻⁸) associated with general cognitive function. Within the novel genetic loci are variants associated with neurodegenerative and neurodevelopmental disorders, physical and psychiatric illnesses, and brain structure. Gene-based analyses find 709 genes associated with general cognitive function. Expression levels across the cortex are associated with general cognitive function. Using polygenic scores, up to 4.3% of variance in general cognitive function is predicted in independent samples. We detect significant genetic overlap between general cognitive function, reaction time, and many health variables including eyesight, hypertension, and longevity. In conclusion we identify novel genetic loci and pathways contributing to the heritability of general cognitive function.
Some individuals have generally higher cognitive function than others. These individual differences are quite persistent across the life course from later childhood onwards. Individuals with higher measured general cognitive function tend to live longer and be less deprived. Retaining general cognitive function is an important aspect of healthy ageing. The population variance in this medically- and socially-important trait has environmental and genetic aetiologies. The details of the genetic contributions are, as yet, poorly understood.

Since the discovery of general cognitive ability (or ‘g’) in 1904, hundreds of studies have replicated the finding that around 40% of the variance in subjects’ scores on a diverse battery of cognitive tests can be accounted for by a single general factor. Some variance is also attributable to individual cognitive domains (e.g., reasoning, memory, processing speed, and spatial ability), and some is attributable to specific cognitive skills associated with individual mental tests. However, all cognitive tests rely to a greater or lesser extent on general cognitive ability for successful execution. Figure 1 illustrates and explains this hierarchical model of cognitive ability differences. Therefore, using a general cognitive function phenotype in a genetically-informative design is supported by the observation that the well-established positive manifold of cognitive tests may be represented by a substantially heritable, higher-order, latent general cognitive function phenotype.

There are two commonly-used routes that are used to obtain general cognitive ability scores for each participant in a sample. First, if all members of a sample have taken the same set of diverse cognitive tests, then a data reduction procedure (such as principal components analysis (PCA) or factor analysis) can be applied. Typically, this finds that all tests load on (i.e., correlate positively with) the first unrotated component, or factor, and scores on this component can be calculated for each person; this gives each person a g score. Second, some mental tests—usually those involving complex mental work, and often those with a variety of item types—have a high g loading. That is, scores on some individual cognitive tests can be used to obtain an acceptable proxy for general cognitive ability. An example of the latter is the Moray House Test of verbal and numerical reasoning, which has a high correlation with a PCA-derived general cognitive function score.

General cognitive function is peerless among human psychological traits in terms of its empirical support and importance for life outcomes. Individuals who have higher cognitive function in childhood and adolescence tend to stay longer in education, gain higher educational qualifications, progress to more professional and better-paid jobs, live healthier lives, and live longer. Individual differences in general cognitive function show phenotypic and genetic stability across most of the life course.

The phenotypic correlation between general cognitive function scores on the same people at age 11 and age 70–80 years is almost 0.7, and remains above 0.5 when age 11 versus age 90 scores are correlated.

Twin studies find that general cognitive function has a heritability of more than 50% from adolescence through adulthood to older age. SNP-based estimates of heritability for general cognitive function are about 20–30%. However, these estimates might increase to about 50% when family-based designs are used to retain the contributions made by rarer SNPs. To date, little of this substantial heritability has been explained, i.e., only a few relevant genetic loci have been discovered (Table 1; Supplementary Fig. 1). As has been found with other highly polygenic traits, a limitation on uncovering relevant genetic loci is sample size; to date, there have been fewer than 100,000 individuals in studies of general cognitive function. The MTAG (multi-trait analysis of genome-wide association studies) method has been used to corral cognitive function and associated traits to expand the number of loci associated with general cognitive function. However, the present study uses only cognitive function phenotypes, and amasses a total sample size of over 300,000.

The present study also tests for genetic contributions to reaction time, and examines its genetic relationship with general cognitive function. Reaction time is both phenotypically and genetically correlated with general cognitive function, and accounts for some of its association with health.

By making these comparisons between general cognitive function and reaction time, we identify regions of the genome that have a shared correlation with general cognitive function and more elementary cognitive tasks.

---

**Fig. 1** The hierarchical model of cognitive function variance. At level 1, individuals differ in specific test and error variance. On all the tests correlate positively. It is found that there are especially strong correlations among the tests of the same domain, so a latent trait at the domain level can be extracted to represent this common variance. It is then found that individuals who do well in one domain also tend to do well in the other domains, so a general cognitive latent trait called $g$ can be extracted. This model allows researchers to partition cognitive performance variance into these different levels. They can then explore the causes and consequences of variance at different levels of cognitive specificity-generality. For example, there are genetic and ageing effects on $g$ and on some specific domains, such as memory and speed of processing. Note that the specific-test-level variance contains variation in the performance of skills that are specific to the individual test and also contains error variance. (Reproduced, with permission, from ref. 13).
Results
General cognitive function phenotypes. The psychometric characteristics of the general cognitive component from each cohort in the CHARGE consortium are shown in Supplementary Note 1. In order to address the fact that different cohorts had applied different cognitive tests, we previously showed that two general cognitive function components extracted from different sets of cognitive tests on the same participants correlate highly \(^{13}\). In order to address the fact that different cohorts had applied different criteria, distributed within 148 loci across all autosomal regions, see Methods section for description of independent SNP associations, and 21,855 at a suggestive level (\(1 \times 10^{-5} \leq P \leq 5 \times 10^{-8}\)) were meta-analyzed using linkage disequilibrium score (LDSC) regression, was estimated at 0.87 (SE = 0.03). This indicates very substantial overlap between the genetic variants associated with cognitive function in these two groups.

SNP-based meta-analyses of cognitive GWASs. We performed an \(N\)-weighted meta-analysis of general cognitive function which included all of the CHARGE, COGENT, and UK Biobank samples. Meta-analysis of the results for the general cognitive function GWASs found 11,600 significant (\(P \leq 5 \times 10^{-8}\)) SNP associations, and 21,855 at a suggestive level (\(1 \times 10^{-5} \leq P \leq 5 \times 10^{-8}\)); see Fig. 2a, Supplementary Fig. 2a, and Supplementary Data 1 and 2. There were 434 independent significant SNPs; see Methods section for description of independent SNP selection criteria, distributed within 148 loci across all autosomal chromosomes. Note that, for consistency, we use the term ‘independent’ here according to the definition that is used in the relevant analysis package. A comparison of these 148 loci with results from the largest previous GWASs of cognitive function \(^{16,17}\), and educational attainment \(^{24}\), and an MTAG analysis of cognitive function GWASs found 11,600 significant genomic risk loci associated with general cognitive function that are potentially functional (Fig. 3a; Supplementary Data 4). See Methods section for further details. Across many of the loci there is clear evidence of functionality including involvement in gene regulation, deleterious SNPs, eQTLs, and regions of open chromatin.

For the 434 independent significant SNPs and tagged SNPs, a summary of previous SNP associations is listed in Supplementary Data 5. They have been associated with many physical (e.g., BMI, height, weight), medical (e.g., lung cancer, Crohn’s disease, blood pressure), and psychiatric (e.g., bipolar disorder, schizophrenia, autism) traits. Of the 58 new loci, we highlight previous associations with schizophrenia (2 loci), Alzheimer’s disease (1 locus), and Parkinson’s disease (1 locus).

We sought to identify independent significant and tagged SNPs within the 148 significant genomic risk loci associated with general cognitive function that are potentially functional (Fig. 3a; Supplementary Data 4). See Methods section for further details. Across many of the loci there is clear evidence of functionality including involvement in gene regulation, deleterious SNPs, eQTLs, and regions of open chromatin.

General cognitive function gene-based and gene-set results. A gene-based association analysis identified 709 genes as significantly associated with general cognitive function (Fig. 2b; Supplementary Fig. 2b; Supplementary Data 6). These 709 genes were compared to gene-based associations from previous studies of general cognitive function and educational attainment \(^{13,16,17,25}\). 418 were replicated in the present study, and 291 were novel. The 291 new gene-based associations are highlighted in Supplementary Data 6. Several of the specific genes associated with general cognitive function are considered in detail in the Discussion, below. Gene-set analysis identified seven significant gene sets associated with general cognitive function: neurogenesis (\(P = 1.57 \times 10^{-6}\)), regulation of nervous system development (\(P = 7.52 \times 10^{-7}\)), neuron projection (\(P = 7.89 \times 10^{-7}\)), positive regulation of nervous system development (\(P = 9.42 \times 10^{-7}\)), neuron differentiation (\(P = 1.68 \times 10^{-6}\)), regulation of cell development (\(P = 1.93 \times 10^{-6}\)), and dendrite (\(P = 3.52 \times 10^{-6}\)) (Supplementary Data 7). Gene-property analysis can show if tissue-specific expression levels are associated with a gene’s association with a phenotype. This analysis indicated a significant association between transcription levels in all brain regions—except the brain spinal cord and cervical c1—and the association with general cognitive function. In addition, expression levels in the pituitary were associated with gene-based association with general cognitive function; these results indicate that the genes with the highest expression levels in these regions were those showing the greatest associations with general cognitive function. (Fig. 3b, c; Supplementary Table 1; Supplementary Data 8). The significance of this relationship was greatest in the cerebellum and the cortex.

**Table 1 Details of GWA studies of general cognitive function to date, including the present study**

<table>
<thead>
<tr>
<th>Author; doi</th>
<th>Year</th>
<th>N</th>
<th>GWAS-sig SNP hits</th>
<th>GWAS-sig gene hits</th>
<th>SNP-based (h^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davies et al. (2011) (^{26})</td>
<td>2011</td>
<td>3511</td>
<td>0</td>
<td>1 gene</td>
<td>0.51 (0.11)</td>
</tr>
<tr>
<td>Lencz et al. (2013) (^{37})</td>
<td>2013</td>
<td>5000</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Benyamin et al. (2014) (^{38})</td>
<td>2014</td>
<td>17,989</td>
<td>0</td>
<td>0</td>
<td>0.46 (0.06)</td>
</tr>
<tr>
<td>Kirkpatrick et al. (2014) (^{39})</td>
<td>2014</td>
<td>7100</td>
<td>0</td>
<td>0</td>
<td>0.35 (0.11)</td>
</tr>
<tr>
<td>Davies et al. (2015) (^{25})</td>
<td>2015</td>
<td>53,945</td>
<td>3 loci (13 SNPs)</td>
<td>1 gene</td>
<td>0.29 (0.05)</td>
</tr>
<tr>
<td>Davies et al. (2016); results for ‘fluid’ test</td>
<td>2016</td>
<td>36,035</td>
<td>3 loci (149 SNPs)</td>
<td>7 loci 17 genes</td>
<td>0.31 (0.02)</td>
</tr>
<tr>
<td>Trampush et al. (2017) (^{44})</td>
<td>2017</td>
<td>35,298</td>
<td>2 loci (7 SNPs)</td>
<td>3 loci 7 genes</td>
<td>0.22 (0.01)</td>
</tr>
<tr>
<td>Snickers et al. (2017) (^{40})</td>
<td>2017</td>
<td>78,308</td>
<td>18 loci (336 SNPs)</td>
<td>47 genes</td>
<td>0.20 (0.01)</td>
</tr>
<tr>
<td>Davies et al. (2018); present study</td>
<td>2018</td>
<td>300,486</td>
<td>148 loci (11,600 SNPs)</td>
<td>709 genes</td>
<td>0.25 (0.006)</td>
</tr>
</tbody>
</table>

For SNP-based heritability, the value from the largest sample is given.
SNP-based heritability of general cognitive function. We estimated the proportion of variance explained by all common SNPs using GCTA-GREML in four of the largest individual samples: English Longitudinal Study of Ageing (ELSA; \( N = 6661, h^2 = 0.12, SE = 0.06 \)), Understanding Society (\( N = 7841, h^2 = 0.17, SE = 0.04 \)), UK Biobank Assessment Centre (\( N = 86,010, h^2 = 0.25, SE = 0.006 \)), and Generation Scotland (\( N = 6,507, h^2 = 0.20, SE = 0.0523 \)) (Table 2). Genetic correlations for general cognitive function amongst these cohorts, estimated using bivariate GCTA-GREML, ranged from \( r_g = 0.88 \) to 1.0 (Table 2). These results indicate that the same genetic variants contribute to phenotypic differences in general cognitive function across each of these three samples. We investigated the genetic contribution to the stability of individual differences in people’s verbal-numerical reasoning, by examining data from those individuals in UK Biobank who completed the test on two occasions (mean time gap = 4.93 years). We found a significant and perfect genetic correlation of \( r_g = 1.0 \) (SE = 0.02).

Polygenic profile scores and genetic correlations. After omitting them from the meta-analysis of GWASs, we created general cognitive function polygenic profile scores in three of the
larger cohorts: ELSA, Generation Scotland, and Understanding Society. The polygenic profile score for general cognitive function explained 2.63% of the variance in ELSA ($\beta = 0.17$, SE = 0.01, $P = 1.70 \times 10^{-51}$), 3.73% in Generation Scotland ($\beta = 0.20$, SE = 0.01, $P = 5.02 \times 10^{-68}$), and 4.31% in Understanding Society ($\beta = 0.22$, SE = 0.01, $P = 6.17 \times 10^{-88}$). Full results for all five thresholds are shown in Supplementary Table 2.

We tested the genetic correlations between general cognitive function and 52 health-related traits. Thirty-six of these health-
related traits were significantly genetically correlated with general cognitive function (Supplementary Data 9). We report significant genetic correlations between general cognitive function and: hypertension ($r_g = -0.15$, SE $= 0.02$), grip strength (right hand: $r_g = 0.09$, SE $= 0.02$), wearing glasses or contact lenses ($r_g = 0.28$, SE $= 0.04$), short-sightedness ($r_g = 0.32$, SE $= 0.03$), long-sightedness ($r_g = -0.21$, SE $= 0.05$), heart attack ($r_g = -0.17$, SE $= 0.03$), angina ($r_g = -0.18$, SE $= 0.03$), lung cancer ($r_g = -0.26$, SE $= 0.05$), and osteoarthritis ($r_g = -0.24$, SE $= 0.04$). We also report a significant genetic correlation with major depressive disorder ($r_g = -0.30$, SE $= 0.04$); this result strengthens previously-reported non-significant correlations of around $-0.10^{16,17}$. We also note the

![Table 2 Genetic correlations and heritability estimates of a general cognitive function component in three United Kingdom cohorts](image)

Below the diagonal, genetic correlations (standard error) of general cognitive function amongst three cohorts are shown: ELSA English Longitudinal Study of Ageing, GS Generation Scotland, US Understanding Society. SNP-based heritability (standard error) estimates appear on the diagonal.

![Fig. 4 Association results for reaction time. SNP-based (a) and gene-based (b) association results in 330,069 individuals. The red line indicates the threshold for genome-wide significance: $P < 5 \times 10^{-8}$ for (a), $P < 2.75 \times 10^{-6}$ for (b); the blue line in (a) indicates the threshold for suggestive significance: $P < 1 \times 10^{-5}$](image)
important genetic association between general cognitive function and longevity \( (r_g = 0.17, \text{SE} = 0.06) \).

**Reaction time results.** GWAS results for mean reaction time uncovered 2022 significant SNPs in 42 independent genomic loci (Fig. 4a; Supplementary Fig. 2c; Supplementary Data 10). Suggestive findings are presented in Supplementary Data 11. Both of the significant loci previously reported for this phenotype were replicated\(^1\). SNPs within the 42 independent genomic loci showed clear evidence of functionality (Fig. 5a; Supplementary

---

**Fig. 5** Functional analyses of reaction time. Analyses include reaction time-associated SNPs, independent significant SNPs, and all SNPs in LD with independent significant SNPs. Functional consequences of SNPs on genes (a) indicated by functional annotation assigned by ANNOVAR. MAGMA gene-property analysis results; results are shown for average expression of 30 general tissue types (b) and 53 specific tissue types (c). The dotted line indicates the Bonferroni-corrected \( \alpha \) level.
Data 12). Using gene-based GWA, a total of 191 genes attained statistical significance (Fig. 4b; Supplementary Fig. 2d; Supplementary Data 13), replicating 18 of the 23 genome-wide significant genes found previously for this phenotype. Gene-set analysis identified no gene sets associated with reaction time (Supplementary Data 14). Gene-property analysis indicated a role for genes expressed in the brain (P = 4.66 x 10^{-15}), with this link between gene transcription levels and gene-based association with reaction time being found across the cortex (Fig. 5b, c; Supplementary Table 3; Supplementary Data 15). Gene transcription levels observed in the pituitary gland were also linked to gene-based associations with differences in reaction time (P = 7.60 x 10^{-9}).

The SNP-based heritability of reaction time was 7.42% (SE = 0.29). It should be noted that this estimate is likely to be an underestimation due to the method used (LD score regression) 39.

Significant overlap was found between the genetic architecture of reaction time and these health outcomes: ADHD, bipolar disorder, schizophrenia, subjective wellbeing, hand grip strength, sleep duration, maternal longevity, hypertension and neurotismism (Supplementary Data 9). The polygenic score for reaction time explained 0.43% of the general cognitive function variance in ELSA (P = 1.42 x 10^{-9}), 0.56% in Generation Scotland (P = 2.49 x 10^{-11}), and 0.26% in Understanding Society (P = 1.50 x 10^{-9}). The full results for all five thresholds can be found in Supplementary Table 2.

We found a genetic correlation (r_g) of 0.247 (P = 1.28 x 10^{-30}) between reaction time and general cognitive function. Overlapping results between the two phenotypes were explored further.

Of the 11,600 genome-wide significant SNPs for general cognitive function, 8269 had a consistent direction of effect with reaction time (sign test, P = 2.2 x 10^{-16}) (Supplementary Data 1). For reaction time, 1070 of the 2022 significant SNPs were consistent for direction of effect with general cognitive function (sign test, P = 0.0071) (Supplementary Data 10). One hundred and sixty SNPs were genome-wide significant for both general cognitive function and reaction time, with 82 consistent for direction of effect (sign test, NS) (Supplementary Data 16). These overlapping genome-wide findings are located within six genomic loci (genomic loci: 13, 15, 19, 28, 69, 133; see Supplementary Data 4 for details of loci); two of these are novel loci for general cognitive function. In the gene-based analyses of both the general cognitive function and reaction time phenotypes, there were 39 overlapping significant genes: 13 of these are newly-identified associations with general cognitive function (Supplementary Data 17).

Discussion

In these meta-analyses of genome-wide association studies for both general cognitive function and reaction time (N = 300,486; N = 330,069, respectively), we make several original contributions. We report 148 genome-wide significant loci for general cognitive function, of which 58 loci have not been reported before. We report 291 gene-based associations for general cognitive function, and 173 for reaction time, which have not been reported previously. Of these genome-wide significant results, six loci and 39 gene-based associations are genome-wide significant for both general cognitive function and reaction time. We are able to predict, using polygenic scoring, up to 4.31 and 0.56% of the general cognitive function variance in an independent sample, for general cognitive function and reaction time polygenic scores, respectively. We present original and updated estimates of genetic correlations with many health traits for both general cognitive function and reaction time. Gene-set analyses identified significant associations for general cognitive function with gene-sets involved in neural and cell development. Significant enrichments were observed with genes expressed in the cerebellum and the brain’s cortex for both general cognitive function and reaction time.

Upon additional exploration of the 58 newly-associated genetic loci, we find that many contain genes that are of further interest. All of the genes discussed below are also genome-wide significant in the general cognitive function gene-based association analysis (P < 2.75 x 10^{-6}; Supplementary Data 6). Significant gene-based associations with general cognitive function have also been previously reported for GATAD2B, SLC39A1, and AUTS2 16,17.

GATAD2B and SLC39A1 are located on chromosome 1; locus 11. Mutations in GATAD2B have been linked to intellectual disability 34. SLC39A1 has been implicated in Alzheimer’s Disease 38. The ATXN1 gene (chromosome 6; locus 60), encodes a protein containing a polyglutamine tract that has previously been associated with spinocerebellar Ataxia 1 29. ATXNL1, ATXNL2, and ATXNL7L2 were also located in significant loci that have previously been associated with cognitive function, intelligence, or educational attainment 16,17,24. The DCDC2 gene (chromosome 6; locus 64) has previously been associated with cortical morphology 30, dyslexia 31, and normal variation in reading and spelling 32, but not with general cognitive function. TTBK1 (chromosome 6; locus 66) encodes a neuron-specific serine/threonine and tyrosine kinase, which regulates phosphorylation of tau 33. Genetic variants in this gene have been associated with Alzheimer’s disease 34. AUTS2 (chromosome 7; locus 72) is implicated in a number of neurological disorders 35. Mutations in CWF19L1 (chromosome 10; locus 91) have been associated with spinocerebellar ataxia and intellectual disability 36. RBFOX1 (chromosome 16; locus 121) encodes a mRNA-spooling factor that interacts with ATXN27, and mutations in this gene lead to neurodevelopmental disorders 37. Locus 131 on chromosome 17, has previously been associated with Smith-Magenis Syndrome 38. The most significantly-associated SNP (P = 2.2 x 10^{-5}) in this locus lies in an intron of the RAI1 gene. RAI1 encodes a protein containing a polymorphic polyglutamine tract that is expressed mainly in neuronal tissues. Variants in the gene are also associated with schizophrenia 39.

Of the seven significant gene sets identified, one was a new finding: ‘positive regulation of nervous system development’. A more detailed description of this gene-set is: ‘any process that activates, maintains or increases the frequency, rate or extent of nervous system development, the origin and formation of nervous tissue’. The remaining six gene-sets showed replication with previous studies of general cognitive function and/or education 16,17,24. Only one, ‘regulation of cell development’, was significant across all four studies 16,17,24. Identification of these genetic sets is consistent with genes associated with cognitive function regulating the generation of cells within the nervous system, including the formation of neuronal dendrites.

A number of not-previously-reported genetic correlations with cognitive function were found here, including with cardiovascular variables. For example, it is already known that there is a phenotypic association between cognitive function in youth and the development of hypertension by age 50 years 41; we found a genetic correlation of −0.15. Other genetic correlations between cardiovascular variables and cognitive function were angina (r_g = −0.18) and heart attack (r_g = −0.17); again, there are known to be phenotypic associations between prior cognitive functioning and various cardiovascular outcomes 11,12.
The genetic correlations between general cognitive function and eyesight were in opposite directions depending on the reported reason for wearing glasses or contact lenses; this was despite an overall positive genetic correlation between general cognitive function and wearing glasses ($r_g = 0.28$). The result for myopia (short-sightedness; $r_g = 0.32$) was consistent with previous evidence of a positive phenotypic $r$ and genetic $r$ correlation between this trait and cognitive function. Less genetic work has investigated the links between hyperopia (long-sightedness) and cognitive function, although our finding, a genetic correlation of $r_g = -0.21$, was consistent with the negative phenotypic association between these variables reported in previous literature.

We have investigated the six regions of the genome identified as having a shared effect between general cognitive function and more elementary cognitive tasks. Locus 13 on chromosome 1 contains the $N_{NNAT}$2 gene. $N_{NNAT}$2 is involved with Wallerian degeneration$^{66-67}$; this is a neurodegenerative process which occurs after axonal injury in both the peripheral and central nervous system. Locus 15 on chromosome 2 contains $E_N{SG}00000271894$, a non-coding RNA gene. $S_{LCA}A10$ and $D_{PP}4$ are located on chromosome 2 (locus 28). Variants in both $S_{LCA}A10$ and $D_{PP}4$ have been linked to schizophrenia$^{38,49}$, hippocampal volume has also been linked to variants in $D_{PP}4$$^{60}$. A variant of $F_OX3$ (chromosome 6; locus 69) has been shown to be associated with longevity in humans$^{51,52}$; it is found in most centenarians across a variety of populations. $M_APT$, $W_N3T$, $C_RHR1$, $K_ANSL1$, and $N_SF$ are located on chromosome 17, locus 133; genetic variants within these genes have been linked to Alzheimer’s disease in $A_POE$ e carriers$^{53}$, Parkinson’s disease$^{54-56}$, neuroticism$^{57}$, infant head circumference$^{58}$, intracranial volume$^{59}$, and subcortical brain region volumes$^{60}$. Research following up the present study’s results could prioritise the genetic loci uncovered herein that are associated with general cognitive function and reaction time (Supplementary Data 16 and 17), as well as those that are also associated with brain-related measures in other large GWASs. Such variants, being associated with multiple cognitive and neurological phenotypes, might help to prioritise potentially causal variants, and help to identify how differences in genotypic sequence are linked to such phenotypic consequences.

We note limitations with the cognitive phenotypes studied. For general cognitive function, phenotypic heterogeneity is a limitation, due to different tests being used in most samples. We also note the small number of cognitive tests being used in the construction of the general cognitive function phenotype in some cohorts. However, we were able to investigate this further by estimating genetic correlations for general cognitive function amongst some of the larger cohorts. These demonstrated strong positive genetic correlations that ranged from $r_g = 0.88-1.0$ (Table 2). There were slight differences in the test questions and the testing environment for the UK Biobank (Supplementary Figure 1). All individuals were aged between 16 and 102 years. Exclusion criteria included clinical stroke (including self-reported stroke) or prevalent dementia (Supplementary Data 18). General cognitive function, unlike height for example, is not measured the same way in all samples. Here, this was mitigated by applying a consistent method of extracting a general cognitive function component from cognitive test data in the cohorts of the CHARGE and COGENT consortia; all individuals were of European ancestry (Supplementary Note 1).

For each of the CHARGE and COGENT cohorts, a general cognitive function component phenotype was constructed from a number of cognitive tasks. Each cohort was required to have tasks that tested at least three different cognitive domains. We avoided taking more than one cognitive test score from any individual cognitive test. Principal component analysis was applied to the cognitive test scores to derive a measure of general cognitive function. Principal component
analyses results for the CHARGE cohorts were checked by one author (JID) to establish the presence of a single component. The scree slope was examined, the percentage of variance accounted for by the first unrotated principal component was noted, and it was checked that all tests had sufficient loading on the first unrotated principal component. Scores on the first unrotated component were used as the cognitive phenotype (general cognitive function). Principal component analyses for the CHARGE cohorts are described in Trampush et al. (pp. 337–338, and Supplementary Table 1)194.

UK Biobank participants were asked 13 multiple-choice questions that assessed verbal-numerical reasoning (VNR: UK Biobank in Second (VNR) cog strength). The VNR score was the number of questions answered correctly in 2 min. Four samples of UK Biobank participants with verbal-numerical reasoning scores were used in the current analyses. The first sample (VNR Assessment Centre) consists of UK Biobank participants who completed the verbal-numerical reasoning test at baseline in assessment centres (n = 107,586). The second UK Biobank sample (VNR T2) consists of participants who did not complete the verbal-numerical reasoning test at baseline but did complete this test at the first repeat assessment visit in assessment centres (n = 11,123). The third UK Biobank sample (VNR MRI) consists of participants who did not complete the verbal-numerical reasoning test at a previous testing occasion but did complete the test at the imaging visit in assessment centres (n = 300,2). The fourth UK Biobank sample (VNR Web-Based) consists of participants who did not complete the verbal-numerical reasoning test at any assessment centre visit, but did complete this test during the web-based cognitive assessment online (n = 46,322). Details of the cognitive phenotypes for all cohorts can be found in Supplementary Note 1.

At the baseline UK Biobank assessment, 496,790 participants completed the reaction time test. Details of the test can be found in Supplementary Note 1. A sample of 330,069 UK Biobank participants with scores on both the reaction time test and genotyping data was used in this study.

Genome-wide association analyses. Genotype-phenotype association analyses were performed within each cohort, using an additional METAL package (with a sample-size weighted model implemented (http://www.sph.umich.edu/csg/abecasis/Metal). The GWAS of reaction time from the UK Biobank sample was performed using the BGENIE v1.2 analysis package (https://jmarchini.org/bgenie/). A linear SNP association model was tested which accounted for genotype uncertainty. Reaction time was adjusted for the familial relationships—site, or familial relationships—were also fitted as required. Cohort-specific quality control procedures, imputation methods, and covariates are described in Supplementary Data 19. Quality control of the cohort-level summary statistics was performed using the EasyQc software210, which implemented the exclusion of SNPs with imputation quality <0.6 and minor allele count <25.

General cognitive function meta-analysis A meta-analysis including all the CHARGE-COGENT and UK Biobank summary results was performed using the METAL package with a sample-size weighted model implemented (http://www.sph.umich.edu/csg/abecasis/Metal).

Reaction time genome-wide association analysis The GWAS of reaction time from the UK Biobank sample was performed using the BGENIE v1.2 analysis package (https://jmarchini.org/bgenie/). A linear SNP association model was tested which accounted for genotype uncertainty. Reaction time was adjusted for the following covariates: age, sex, genotyping batch, genotyping array, assessment centre, and 40 principal components.

Genomic risk loci characterization using FUMA. Genomic risk loci were defined from the SNP-based association results, using Functional Mapping and Annotation of genetic associations (FUMA)211. Firstly, independent significant SNPs were identified using the SNP2GENE function and defined as SNPs with a P-value of ≤5 × 10^{-8} and independent of other genome wide significant SNPs at r2 < 0.6. Using these independent significant SNPs, tagged SNPs were used to be in subsequent annotations were identified as all SNPs that had a MAF ≥0.0005 and were in LD of r2 ≥ 0.6 with at least one of the independent significant SNPs. These tagged SNPs included those from the 1000 genomes reference panel and need not have been included in the GWAS performed in the current study. Genomic risk loci that were 250 kb or closer were merged into a single locus. Lead SNPs were also identified using the independent significant SNPs and were defined as those that were independent from each other at r2 < 0.1.

Comparison with previous findings. Previous evidence of association for each of the 148 genetic loci identified herein as being associated with general cognitive function was sought in the largest published GWASs of general cognitive function212,213 and education214. We performed look-back meta-analyses at tagged SNPs (r2 > 0.6) within each locus, including all 1000 genomes SNPs, and classed any tagged SNP previously reported as genome-wide significant, as replication. Details of these findings are presented in Supplementary Data 3.

Gene-based analysis implemented in FUMA. Gene-based analysis has been shown to increase the power to detect genotype-phenotype association because the multiple testing burden is reduced, and the effect of multiple SNPs is combined together215. Gene-based analysis was conducted using MAGMA216. The test carried out using MAGMA, as implemented in FUMA, was the default SNP-wise test using the mean χ2 statistic derived on a per gene basis. SNPs were mapped to genes based on the nearest gene from the chromosomal location. All SNPs that were located within the gene body were used to derive a P-value describing the association found with general cognitive function and reaction time. The SNP-wise model from MAGMA was used and the NCBI build 37 was used to determine the location and boundaries of 18,199 autosomal genomic linkage disequilibrium blocks between each SNP and a variant in the most common genotypes used in the 1000 genomes phase 3 release217. A Bonferroni correction was applied to control for multiple testing; the genome-wide significance threshold was P < 2.75 × 10^{-8}.

Estimation of SNP-based heritability. The proportion of variance explained by all common SNPs was estimated using univariate GCTA-GREML analyses218 in four of the different UK Biobank participants’ verbal-numerical reasoning test (VNR: UK Biobank in Second (VNR) cog strength). This indicates that only 4.75% of the in

Univariate Linkage Disequilibrium Score regression. Univariate LDSC regression was performed on the summary statistics from the GWAS on general cognitive function and reaction time. The heritability Z-score provides a measure of the polygenic signal found in each data set. Values greater than four indicate that the data are suitable for use with bivariate LDSC regression219. The mean χ2 statistic indicates the inflation of the GWAS test statistics that, under the null hypothesis of no genetic association (i.e., no inflation of test statistics), would be one. An inflation of test statistics can indicate population stratification, cryptic relatedness, or the presence of many alleles each with a small effect. The intercept of the LDSC regression can detect the difference between inflation due to stratification and cryptic relatedness, and the inflation due to a polygenic signal. This is because the inflation in test statistics attributable to stratification, drift, and cryptic relatedness will not correlate with LD, whereas inflation due to polygenicity will. The LDSC regression intercept, therefore, captures the inflation in the χ2 statistics that is not due to stratification or other confounds.

For each GWAS, an LD regression was carried out by regressing the GWAS test statistics (χ2) on to each SNP’s LDP score, which is the sum of squared correlations between the minor allele frequency count of a SNP with the minor allele frequency count of every other SNP. The LD regression intercept, therefore, captures the inflation due to the slope, and a means to detect residual confounders using the intercept. For general cognitive function, we report a LD score regression intercept of 1.058 (SE = 0.011) and a ratio of 0.0659; this indicates that only 6.6% of the inflation observed can be ascribed to causes other than a polygenic signal. For reaction time, we report an LD score regression intercept of 1.02 (SE = 0.009) and a ratio 0.0475; this indicates that only 4.75% of the inflation observed can be ascribed to causes other than a polygenic signal.

LD scores and weights were downloaded from (http://www.broadinstitute.org/mpg/ldscore) for use with European populations. A minor allele frequency cut-off of ≥0.1 and an imputation quality score of ≥0.9 were applied to the GWAS summary statistics. Following this, SNPs were retained if they were found in HapMap 3 with MAF ≥0.05 (MACH reference sample). For those SNPs, indels and structural variants were removed along with strand ambiguous variants. SNPs whose alleles did not match those in the 1000 Genomes were also removed. As the presence of outliers can increase the standard error in LDSC score regression220 and so SNPs where χ2 > 80 were also removed.

Genetic correlations. Genetic correlations were estimated using two methods, bivariate GCTA-GREML212 and LDSC221. Bivariate GCTA was used to calculate genetic correlations between phenotypes and cohorts where the genotyping data were available. This method was used to calculate the genetic correlations between different cohorts for the general cognitive function phenotype. It was also employed to investigate the genetic contribution to the stability of the same UK Biobank participants’ verbal-numerical reasoning test scores in the assessment centre and then in web-based, online testing. In cases where only GWA summary results were available, bivariate LDSC was employed to calculate genetic correlations between two traits. This was used to estimate the degree of overlap between polygenic architecture of the traits. Bivariate LDSC regression was used to estimate genetic correlations between general cognitive function, reaction time, and the following health outcomes: ADHD, age at menarche, age at menopause, Alzheimer’s disease, anorexia nervosa, bipolar disorder, BMI, bone density femoral neck, bone density lumbar spine, coronary artery disease, HBAlC, HDL cholesterol, hippocampal volume, LDL cholesterol, lung cancer, major depression, neuroticism, schizophrenia, smoking status, trycglycerides, type 2 diabetes, waist-hip ratio, autism spectrum disorder, birth weight, depressive symptoms, hypetension, pulse wave arterial stiffness, angina, heart attack, parental education, forced expiratory volume in 1 second (FEV1), general health perception, happiness, health satisfaction, heel bone mineral density, osteoarthritis, overall health rating, wearing of glasses or contact lenses, long-sightedness, short-sight

DOI: 10.1038/s41467-018-03462-x | www.nature.com/naturecommunications
Alzheimer’s disease, a 500-kb region surrounding APOE was excluded and the analysis re-run (Alzheimer’s disease (500 kb)). Supplementary Data 20 provides further details on the sources of the GWAS summary statistics.

**Polygenic prediction.** Polygenic profile score analyses were used to predict cognitive test performance in Generation Scotland, the English Longitudinal Study of Ageing, and Understanding Society. Polygenic profiles were created in PRSice using results of a general cognitive function meta-analysis that excluded the Generation Scotland, the English Longitudinal Study of Ageing, and Understanding Society cohorts. Polygenic profiles were also created in these cohorts based on the UK Biobank GWA reaction time results. SNPs with a MAF < 0.01 were removed prior to creating the polygenic profiles. Clumping was used to obtain SNPs in linkage disequilibrium with an r² > 0.25 within a 250 kb window. Polygenic profile scores were created at P-value thresholds of 0.01, 0.05, 0.1, 0.5, and 1 (all SNPs), based on the significance of the association in the general cognitive function and reaction time GWAS. Linear regression models were used to examine the associations between the polygenic profile and cognitive ability in GS, ELSA, and US, adjusting for age at measurement, sex, and the first 10 (GS), 15 (ELSA), and 20 (US) genetic principal components to adjust for population stratification. The false discovery rate (FDR) method was used to correct for multiple testing across the polygenic profiles at all five thresholds.

**Functional annotation implemented in FUMA.** The independent significant SNPs and those in LD with the independent significant SNPs were annotated for functional consequences using functions using ANNOVAR and the Ensembl genes build 85. A CADD score37,38, RegulomeDB score41, and 15-core chromatin state42–44 were obtained for each SNP. eQTL information was obtained from the following databases: GTEx (http://www.gtexportal.org/home/), BRAINEAC (http://www.braineac.org/), Blood eQTL Browser (http://genenetwork.nl/bloodeqtlbrowser/), and BIOS QTL browser (http://genenetwork.nl/bioseqtlbrowser/). Functionally-annotated SNPs were then mapped to genes based on physical position on the genome, eQTL associations (all tissues) and chromatin interaction mapping (all tissues). Intergenic SNPs were mapped to the two closest up- and down-stream genes which can result in their being assigned to multiple genes.

**Gene-set analysis implemented in FUMA.** In order to test whether the polygenic signal measured in each of the GWASs clustered in specific biological pathways, a competitive gene-set analysis was performed. Gene-set analysis was conducted in MAGMA using competitive testing, which examines if genes within the gene set are more strongly associated with each of the cognitive phenotypes than other genes. Such competitive tests have been shown to control for Type I error rate as well as facilitating an understanding of the underlying biology of cognitive differences39,40. A total of 10,891 gene-sets (sourced from Gene Ontology, Reactome, and SigDB45) were examined for enrichment of general cognitive function and reaction time. A Bonferroni correction was applied to control for the multiple tests performed on the 10,891 gene sets available for analysis.

**Gene-property analysis implemented in FUMA.** A gene-property analysis was conducted using MAGMA in order to indicate the role of particular tissue types that influence differences in general cognitive function and reaction time. The goal of this analysis was to test if, in 30 broad tissue types and 53 specific tissues, tissue-specific differential expression levels were predictive of the association of a gene with general cognitive function and reaction time. Tissue types were taken from the GTEx v6 RNA-seq database53,54 with expression values being log2 transformed with a pseudocount of 1 after winsorising at 50, with the average expression value being taken from each tissue. Multiple testing was controlled for using a Bonferroni correction.

**Data availability.** The GWAS summary results for all significant and suggestive SNPs for general cognitive function and reaction time are available in Supplementary Data 1, 2, 10 and 11. The full GWAS summary results for Reaction Time are available to download here: http://www.cccse.ed.ac.uk/node/335. Access to the full GWAS summary results for general cognitive function can be requested by application to the chairs of the CHARGE and COGENT consortia.

Received: 31 August 2017 Accepted: 23 April 2018
Published online: 29 May 2018

**References**

and Mental Health, University of Toronto, Toronto M5T 1L8, Canada. 167Department of Public Health, University of Helsinki, Helsinki 00014, Finland. 168Department of Radiology, Erasmus MC, Rotterdam 3015, The Netherlands. 169Department of Psychiatry and Behavioral Sciences, Faculty of Medicine, University of Crete, Heraklion GR-71003, Greece. 160Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston 77030, TX, USA. 161Human Genome Sequencing Center, Baylor College of Medicine, Houston 77030-3411 TX, USA. 162Neuropsychiatric Genetics Research Group, Department of Psychiatry and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin D02 AY89, Ireland. 163Center for Translational and Systems Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York 10032 NY, USA. 164Program in Medical and Population Genetics, Broad Institute, Cambridge 02142 MA, USA. 165Department of Neurology, University Hospital of Bordeaux, Bordeaux 33000, France. 166Department of Global Health, University of Washington, Seattle 98104 WA, USA. 167Department of Cardiology, Leiden University Medical Center, Leiden 2333, The Netherlands. 168Institute for Behavioral Genetics, University of Colorado, Boulder 80309 CO, USA. 169Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK. 166Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki 00014, Finland. 167Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Manchester Academic Health Science Centre, and Manchester Medical School, Institute of Brain, Behaviour, and Mental Health, University of Manchester, Manchester M13 9PL, UK. 168Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland. 169School of Medicine and Public Health, University of Newcastle, New South Wales 2308, Australia. 170Neuroscience Research Australia, Sydney 2031, Australia. 171Faculty of Medicine, University of New South Wales, Sydney 2052, Australia. 172Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK. 173Division of Nephrology and Hypertension, Department of Internal Medicine, Mayo Clinic, Rochester, MN 55905, USA. 174Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany. 175Day Clinic for Cognitive Neurology, University Hospital Leipzig, Leipzig 04103, Germany. 176Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks 11004 NY, USA. 177Department of Psychiatry, Hofstra Northwell School of Medicine, Hempstead 11549 NY, USA. 178Division of Psychiatry, University of Edinburgh, Edinburgh EH10 5HF, UK. 179MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton SO16 6YJ, UK. 180Glenn Biggs Institute for Alzheimer’s and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio 78229 TX, USA. These authors contributed equally: Gail Davies, Max Lam. These authors jointly supervised this work: Todd Lencz, Ian J. Deary.