

SCIENTIFIC REPORTS



OPEN

Enrichment of genetic markers of recent human evolution in educational and cognitive traits

Saurabh Srinivasan^{1,2}, Francesco Bettella^{1,2}, Oleksandr Frei^{1,2}, W. David Hill^{3,4}, Yunpeng Wang^{1,2}, Aree Witoelar^{1,2}, Andrew J. Schork⁸, Wesley K. Thompson^{7,8}, Gail Davies^{3,4}, Rahul S. Desikan⁹, Ian J. Deary^{3,4}, Ingrid Melle^{1,2}, Torill Ueland^{1,2}, Anders M. Dale^{5,6,10}, Srdjan Djurovic^{11,12}, Olav B. Smeland^{1,2} & Ole A. Andreassen^{1,2}

Higher cognitive functions are regarded as one of the main distinctive traits of humans. Evidence for the cognitive evolution of human beings is mainly based on fossil records of an expanding cranium and an increasing complexity of material culture artefacts. However, the molecular genetic factors involved in the evolution are still relatively unexplored. Here, we investigated whether genomic regions that underwent positive selection in humans after divergence from Neanderthals are enriched for genetic association with phenotypes related to cognitive functions. We used genome wide association data from a study of college completion (N = 111,114), one of educational attainment (N = 293,623) and two different studies of general cognitive ability (N = 269,867 and 53,949). We found nominally significant polygenic enrichment of associations with college completion (p = 0.025), educational attainment (p = 0.043) and general cognitive ability (p = 0.015 and 0.025, respectively), suggesting that variants influencing these phenotypes are more prevalent in evolutionarily salient regions. The enrichment remained significant after controlling for other known genetic enrichment factors, and for affiliation to genes highly expressed in the brain. These findings support the notion that phenotypes related to higher order cognitive skills typical of humans have a recent genetic component that originated after the separation of the human and Neanderthal lineages.

The evolution of cognitive function and brain development is regarded as the result of a complex interplay of nature and nurture, where development seems to be driven by genes and shaped by environment¹. Modern humans have highly complex brains, capable of processing vast information and solving abstract problems. In addition, humans have enhanced cognitive functioning, especially in the domains of cooperation, egalitarianism, theory of mind, language and culture, and achieved new modes of thinking and reasoning that seem to have greatly increased their ability to flourish as a species². Archaeological and fossil records provide evidence suggesting anatomical and morphological evolution of humans and their predecessors^{3,4}. However, we have no direct evidence for the evolution of higher cognitive functions and must rely on cultural artefacts that indirectly suggest behavioural changes⁵.

Several lines of evidence support the idea that humans have developed complex language⁵, executive functioning, and abstract thinking in the process of evolution⁶. Language, and the thoughts that it expresses, arguably

¹NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ²Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. ³Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. ⁴Department of Psychology, University of Edinburgh, Edinburgh, UK. ⁵Multimodal Imaging Laboratory, University of California at San Diego, La Jolla, CA, USA. ⁶Center for Human Development, University of California at San Diego, La Jolla, CA, USA. ⁷Institute of Biological Psychiatry, Mental Health Center St. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. ⁸Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. ⁹Neuroradiology Section, Department of Radiology and Biomedical Imaging, University of California at San Francisco, San Francisco, CA, USA. ¹⁰Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. ¹¹Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. ¹²NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway. Correspondence and requests for materials should be addressed to O.A.A. (email: o.a.andreassen@medisin.uio.no)

constitute the most distinctive features of the modern human mind⁷. With specialized knowledge and means to communicate, humans seem to be able to create efficient tools, codify knowledge, make rules and organize society⁸. While we do not know how Neanderthals would have performed in cognitive tests, cognitive studies comparing humans and chimpanzees have found different patterns of performance across cognitive domains suggesting a role of evolution in specific higher cognitive functions⁹. Anatomically, we observe increased general cephalization, which is said to be associated with greater behavioural complexity, and enlargement and specialization of brain regions adapting to various sensory uses over the course of evolution¹⁰. Language and social skills are said to have evolved together; better language skills would be required to facilitate better cooperation while hunting and foraging as a group. Meanwhile, living in social groups may have helped in the development of the neocortex, or the human “social brain”¹¹.

Evolutionary psychologists discuss whether our cognitive processes and environment co-evolved incrementally, instead of undergoing a dramatic change^{1,12}. Recent studies suggest that Neanderthals could have had some ability to express themselves artistically and some sort of proto culture or religion¹³. This is in line with the notion that human intelligence, while not completely innate, is shaped by natural selection and evolutionary processes which helped the species adapt to the environment. Genes and cultures are suggested to co-evolve¹⁴, as in the case of the human ability to learn as a group, which seems to set them apart from our primate relatives and shaped human specific phenotypes. Social learning could help humans to acquire new skills faster and to enhance them¹⁵. Cognitive function and academic achievement, while influenced by genes, probably need the right environment to achieve full potential¹⁶. Educational attainment (or the ability to complete college) is governed not only by many socio-economic, political and cultural factors^{17,18} but also by genetic factors that were estimated to account for approximately 20% of the phenotypic variance¹⁹. In twin studies, the heritability of general cognitive function was estimated to be 50–60%^{20,21}. Educational attainment is also related to cognitive performance with a high genetic correlation^{22,23}. Cognitive function is in turn often associated with neuropsychiatric phenotypes, with which it shows some genetic overlap^{16,24–27}. Finally, several cognitive traits, such as attention, memory and reasoning, are important to succeed in academics.

Neanderthals, considered a sister group of modern humans, are suggested to have split from the human lineage between 500,000 to 750,000 years ago^{28–30}. As the closest living relatives to humans, chimpanzees are often used as a reference point for ancestral alleles. Chimpanzees split from humans approximately 6.3 million years ago³¹. The chimpanzee genome was sequenced in 2005, and aligns with 96 to 98% of the human genome^{32,33} depending on the exact criterion utilized for sequence alignment. Recent studies have found that variants associated with cognitive function are enriched in regions of the genome that are evolutionarily conserved in mammals³⁴. However, these do not provide a human specific time frame. The availability of Neanderthal and chimpanzee genomes makes it possible to determine when in the course of human evolution genomic regions underwent selective pressures by cross-referencing the chimpanzee, human and Neanderthal sequences. This provides a rare opportunity to gain novel insight into the evolutionary processes in humans.

We have previously shown how these available genomes can be used to detail the evolution of schizophrenia, employing the original version²⁹ of the Neanderthal selective sweep score³⁵. Here, we applied the same polygenic enrichment approach in conjunction with a more recent and comprehensive post-Neanderthal selective sweep index³⁶ to study the evolutionary aspects of educational attainment and cognitive function, and determine enrichment in genomic regions that may have undergone recent positive selection in humans. To this end, we analysed genome-wide association (GWAS) summary statistics for two measures of educational attainment, college completion (College)³⁷ and years of educational attainment (EduYears)³⁸, and two measures of fluid intelligence from two studies of general cognitive ability (GCA)^{39,40}. The GCA metric is designed to capture around 40–50% of the variance across diverse cognitive abilities, irrespective of the specific tests used to construct it^{41,42}. We hypothesized that higher cognitive functions are a product of human evolution in line with previous theories⁴³. We compared the cognitive phenotypes to height and body mass index (BMI), two human traits with GWASs of similar size, to assess the specificity of this enrichment.

Results

Utilizing a post-Neanderthal selective sweep (PNSS) index³⁶, we assessed the effect of a variant’s affiliation to the selectively swept regions of the genome on traits related to cognition: College completion (College) (N = 111,114), education attainment (EduYears) (N = 293,623) and two measures of general cognitive ability (GCA), GCA1 (N = 269,867) and GCA2 (N = 53,949)). The PNSS index defines regions that have undergone positive selection in humans after the separation of the human and Neanderthal lineages. We specifically investigated association enrichment, visualized as an upward deflection in fold enrichment plots and a leftward deflection in Q-Q plots (for details, see Methods). The fold enrichment plots (Fig. 1) and conditional Q-Q plots (Supplementary Fig. S1) suggest that the genetic variants in regions that may have undergone positive selection in humans, i.e. the human divergent (HD) regions, are markedly enriched of associations with College, EduYears and GCA1, and to a lesser extent with GCA2.

To test the significance of the enrichment, we conducted a stratified LD score regression analysis⁴⁴. The LD score regression method provides an estimate of the fold enrichment associated with these evolutionary regions, and thus an estimate of enrichment⁴⁵. Specifically, we found significant enrichment for EduYears (fold enrichment = 2.00 vs. expected, $p = 0.044$), GCA2 (3.48, $p = 0.023$), College (2.33, $p = 0.026$) and GCA1 (2.12, $p = 0.024$) (Table 1). The LD score regression analysis also provides a regression coefficient controlling for affiliation to generic functional categories such as intron, exon, 3’UTR, 5’UTR, and to brain genes in particular (Supplementary Tables S1–S2). Our analysis indicates that affiliation to HD regions significantly contributes to the LD score effect in College ($p = 0.019$), GCA2 ($p = 0.020$), GCA1 ($p = 0.029$) and EduYears ($p = 0.046$) after controlling for the other covariates.

Phenotype	Enrichment	
	Fold \pm s.e.m	p-value
BMI	1.38 \pm 0.374	3.12×10^{-1}
College	2.33 \pm 0.605	$2.63 \times 10^{-2*}$
EduYears	2.00 \pm 0.605	$4.48 \times 10^{-2*}$
GCA 1	2.15 \pm 0.579	$2.45 \times 10^{-2*}$
GCA 2	3.48 \pm 1.099	$2.36 \times 10^{-2*}$
Height	2.10 \pm 0.607	7.17×10^{-2}

Table 1. Post-Neanderthal selective sweep enrichment. *Nominally significant. Stratified enrichment analysis for genetic variants in regions that possibly underwent a selective sweep after divergence from Neanderthal. Cognitive measures: college completion (College), educational attainment (EduYears), general cognitive ability (GCA1 and 2); Anthropometric measures: body mass index (BMI) and Height. The table shows the LD score regression model enrichment test statistics: the enrichment of associations among variants in swept regions compared to other variants (Fold), and the corresponding p-values.

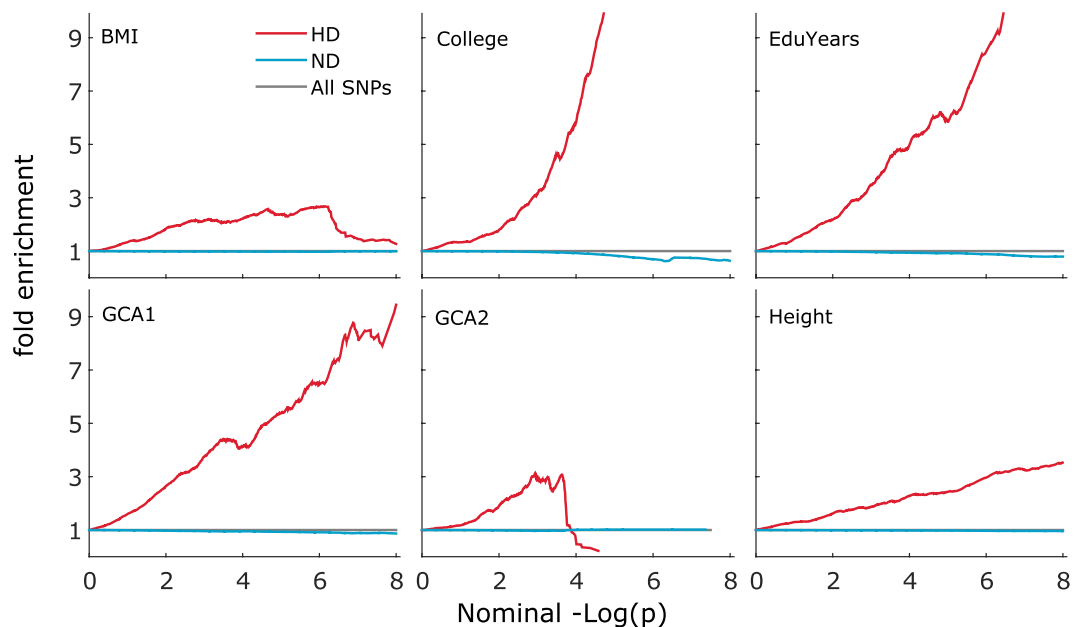


Figure 1. Fold enrichment plots of phenotypes stratified according to Neanderthal selective sweep score. Shown are the fold enrichment plots of GWAS summary statistics p -values for body mass index (BMI), college completion (College), educational attainment (EduYears), general cognitive ability (GCA1 and 2) and height stratified based on Neanderthal selective sweep (NSS) score. The human divergent (HD) stratum comprises single nucleotide polymorphisms (SNPs) with negative NSS scores. The non-divergent (ND) stratum comprises all SNPs with positive NSS scores. HD SNPs show some upward deflection from ND and all SNPs. This signifies a comparatively higher proportion of low p -values among HD SNPs.

To further detail the specificity of the enrichment, we used summary statistics from height and BMI GWASs, which involved sample sizes comparable to or larger than the ones available to the cognitive GWASs. While height and BMI have been associated with some evolutionary pressure in humans^{46,47}, we expect them to show a less pronounced enrichment than the cognitive phenotypes since they are less likely to be human specific compared to cognitive function.

Our fold enrichment plots (Fig. 1) suggest the presence of some enrichment in Height and BMI among variants in HD regions. The deflections are more consistent than observed for GCA2 but less pronounced than seen for College, EduYears and GCA1. The fold enrichment test statistics (BMI: 1.38, $p = 0.312$; Height: 2.10, $p = 0.072$) are non-significant, as are the regression coefficients (BMI $p = 0.252$ and Height $p = 0.080$). Upon meta-analysing the stratified LD-score enrichment test statistics for anthropometric and cognitive traits, a significant effect is found for the latter (Fisher-combined test $p = 4.00 \times 10^{-4}$) but not for the former (Fisher-combined test $p = 0.107$).

Given the importance of genes involved in brain function for the phenotypes of interest, and their evolutionary relevance, we performed additional analyses targeting genes with high expression levels in the brain (brain genes) (Fig. 2). Brain genes in HD regions show more pronounced association enrichment than any brain genes or any SNPs in HD regions in the fold enrichment plots. The stratified LD-score regression analysis suggests brain genes

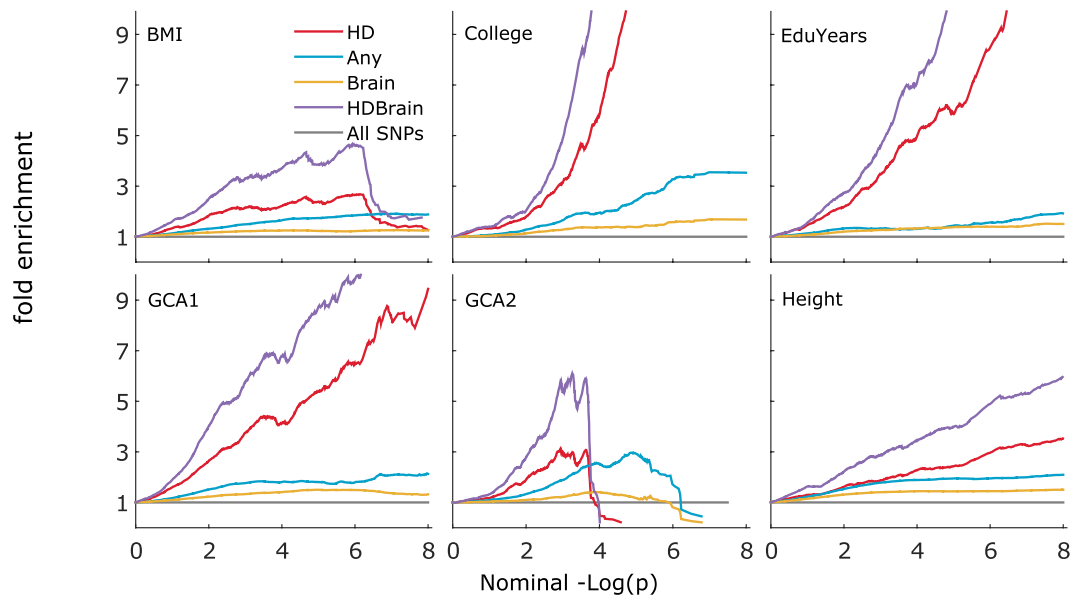


Figure 2. Enrichment of associations with various traits stratified according to their affiliation to Post-Neanderthal selective sweep regions and brain genes. Fold enrichment plots for body mass index (BMI), college completion (College), educational attainment (EduYears), general cognitive ability (GCA1 and 2) and height. Plots are shown for variants annotated to human divergent regions (HD), variants annotated to generic genes (Any), variants annotated to genes with high expression in the brain (Brain), variants in HD regions annotated to genes with high expression in brain (HDBrain), and all variants (All SNPs). The HDBrain category is enriched (upward deflection) compared with the other categories (i.e., presents a higher incidence of associations [lower p-values]).

in the HD regions (HD Brain) to be more enriched than any SNPs in HD regions (HD) in GCA2 (4.75 vs. 3.48) but not in College, EduYears and GCA1 (College = 2.16 vs. 2.33, EduYears = 1.83 vs. 2.00, GCA1 = 1.68 vs. 2.12). However, the numbers of SNPs in these strata are very low and the stratified LD-score regression analysis does not give conclusive results (Supplementary Table S2).

Discussion

Applying our Neanderthal polygenic enrichment approach to recent large GWAS data on cognitive traits^{37–39,48}, we investigated the hypothesis that higher cognitive functions in humans have a recent evolutionary component. We assessed the extent to which the cognitive phenotypes, College, EduYears, and GCA are affected by genetic variation in regions of the human genome that may have undergone selective sweeps since divergence from the Neanderthals and found nominally significant enrichment with all cognitive traits. Together, these findings lend support to the hypothesis that higher cognitive traits typical of humans have a component that originated after the separation of the human and Neanderthal lineages, in line with previous theories⁴³.

The fold enrichment and Q-Q plots for the three cognitive phenotypes showed various degrees of deflection in the HD stratum (Fig. 1 and Supplementary Fig. S1). The fold enrichment statistics confirmed consistently significant enrichments for College, EduYears, and two independent data sets on GCA. The phenotypes summarized under GCA measure fluid intelligence and capture the shared variance across cognitive traits, irrespective of the tests applied. GCA is also phenotypically and genetically correlated with educational attainment^{37,49}.

The meta-analysis of the stratified LD-score enrichment tests statistics suggests that the enrichments detected here are somewhat specific to cognitive phenotypes. In a previous study, we found Height and BMI to be significantly enriched in HD regions. However, those results were obtained using the original Neanderthal selective sweep score²⁹ and a different statistical method³⁵.

The current study sheds light on the evolutionary architecture of human cognitive traits. These phenotypes are complex and involve many genetic and environmental factors. Our results suggest an underlying evolutionary factor in the genetics of educational attainment and cognitive function. Recent studies have indeed found that genes associated with educational attainment are under selective pressure⁵⁰. Behavioural studies suggest that humans have evolved to outperform other primates in certain higher cognitive abilities, such as learning through imitation and from the mistakes of others, and are also quicker to learn unfamiliar tasks^{51–55}. Paleontological or archaeological findings indicate the increasing complexity of cranial anatomy and material culture during human evolution^{3,4}. Our results constitute further evidence that these traits are related to the evolution of modern humans since the divergence from Neanderthals about 500,000 to 750,000 years ago.

The PNSS region affiliation score is an index of possible positive selection in regions of the human genome after divergence from Neanderthals³⁶, but provides no direct experimental evidence that positive selection occurred at those sites. Nonetheless, the Neanderthal selective sweep score compiled on the first edition of the Neanderthal sequence by Green²⁹, and the subsequent post Neanderthal swept regions identified by Prüfer³⁶,

proved to be more sensitive than other genomic proxies of positive selection like human accelerated regions and segmental duplications used in our previous studies⁵⁶. The Neanderthal sweep scores are also more specific than other measures of negative selection in mammals³⁴ in detecting association enrichment.

In the evolution of complex traits, polygenic selection involving subtle shifts of allele frequencies at many loci simultaneously may have been more common than major shifts induced by strong forces⁵⁷. Thus, selection acting simultaneously on many standing variants could have been the more efficient mechanism for phenotypic adaptation^{58,59}. The PNSS may therefore not be in the best position to account for genetic drift or other neutral selection forces. However, the distinction between weak and strong selection is not clear. The methods applied in our analyses have been useful in studying several aspects related to polygenic factors in complex human phenotypes before^{35,56,60–63}. Although the evolutionary proxy, the PNSS, is designed to detect positive selection sweeps rather than polygenic adaptation, the present results are in line with previous theories positing that higher cognitive functions are a product of human evolution⁴³.

While the combined analysis of data from the large cognitive and educational GWASs and the Neanderthal genome sequence is unique to this study, it entails some limitations. The power of the different GWASs depends on their size and on the genetic architecture and polygenicity⁶⁴ of the traits analysed. Hence, the differences in detected enrichments are not strictly comparable to one another. The evolutionary enrichment observed for cognitive phenotypes could be influenced by affiliation to brain genes and genetic functional elements such as introns, exons, 5'UTR and 3'UTR. Despite control for these factors in the analyses, they may be overrepresented in the information-rich portions of the Neanderthal DNA that could be reconstructed. Also, while LD was properly accounted for in the statistical analyses, the enrichment in the plots could be influenced by LD-tagging effects. This could be the reason for the discrepancy observed between the visual and the statistical assessments. Given the complexity of cognitive phenotypes, the observed enrichment may be confounded by other factors with a potential evolutionary component. For example, educational attainment is known to be influenced by other pathologies such as ADHD⁶⁵ and genes associated with ADHD may have some evolutionary advantage as well. Future replication of this result is warranted using summary statistics from larger educational attainment GWAS⁶⁶. Our previous experience suggests that increases in sample size tend to enhance polygenic enrichments³⁵.

In conclusion, we demonstrate that the genetic architectures of two measures of educational attainment and two versions of GCA are enriched for genomic regions that were likely subjected to selective sweeps since divergence from Neanderthals. This suggests that some genetic components of higher cognitive functions in humans are driven by more recent evolutionary processes. These findings should be confirmed in independent studies that could also identify the specific genetic variants involved, to inform the biological underpinnings of human cognitive function.

Materials and Methods

Samples. Educational attainment has a well-documented health-education gradient as well as phenotypic and genetic relation to cognitive functioning⁶⁷, and is influenced by environmental and genetic factors^{19,68}. We obtained summary statistics for about ten million single nucleotide polymorphisms (SNPs) from a GWAS of educational attainment (EduYears)³⁸ (sample N = 328,917 Caucasian individuals from North America, Western Europe and Australia), as well as from UK Biobank GWASs of college or university degree (College)³⁷ (sample N = 111,114) and of general cognitive ability (GCA1)⁴⁰ (sample N = 269,867). The data on GCA1 was based on 269,867 individuals drawn from 14 cohorts, primarily consisting of data from the UK Biobank (sample N = 195,653) and the Cognitive Genomics Consortium (sample N = 35,289). We also obtained summary statistics for the same set of SNPs from a similar GWAS of GCA (i.e., GCA2) in middle and older age by the CHARGE consortium, which included a total of 53,949 individuals³⁹. Finally, we used Height (sample N = 183,727) and BMI (sample N = 339,224) GWAS summary statistics from the GIANT consortium study^{69,70} as control sets.

Analytical Approach. We employed genetic enrichment methods recently developed to uncover more of the genetic architecture of complex traits^{60–62,71}. Specifically, we investigated the enrichment of associations concurrent with the evolutionary affiliations in a covariate-modulated statistical framework⁶¹. We investigated whether variants in evolutionarily salient regions or tagging other variants therein, are more likely associated with measures of education attainment and general cognitive function, as well as with other control phenotypes. The visual displays of enrichment were produced with MATLAB (www.mathworks.com/products/matlab). The enrichment test statistics were computed using the LD-score package⁴⁴.

Cognitive phenotypes. College completion (College) measures the highest level of educational qualification achieved³⁷ while educational attainment (EduYears)³⁸ measures the years of completed schooling. These are used as a proxy for intelligence and they show high genetic correlation⁴⁸. The general cognitive ability is not a specific cognitive skill but a measure of various fluid cognitive ability tests. The measures used in GCA1 and 2 were constructed using the first un-rotated component extracted from a principal component analysis of the individual cognitive test scores that measured general fluid cognitive functions³⁹. These measures of fluid intelligence correlate highly with general cognitive ability^{40,48,72}. The scores used by these two independent phenotypes capture the shared variance across cognitive test batteries measuring fluid cognitive functions, and explain around 40–50% of the variation across cognitive domains. Further details of the tests administered can be found in the original publications^{37,39,40}.

Post-Neanderthal selective sweep regions. The index of the post-Neanderthal selective sweep (PNSS) regions was obtained from the work of Prüfer *et al.*³⁶ and is downloadable from, <http://cdna.eva.mpg.de/neanderthal/>. The authors used a hidden Markov model to identify regions in Neanderthals that differ from modern humans³⁶. Neanderthal and Denisovan genes were used to identify regions that differed from the representative

modern human population in the 1000 genomes project variants. The identified regions were assigned a score based on genetic lengths and a cut off was assigned for regions that were most likely to have undergone positive selection in humans.

We assigned all SNPs a value of 0 or 1 based on whether these fell outside or inside the regions of recent positive selection in humans, respectively.

Confounding/mediating effects. We controlled for the following factors while assessing the evolutionary enrichment of cognition/education associations:

Brain genes. We used the protein atlas (<http://www.proteinatlas.org/humanproteome/brain>) to select all genes that are expressed specifically in the brain of *Homo sapiens*. We identified a total of 4915 genes by filtering for genes that have high expression levels in brain. The 1000 Genomes Project SNPs were then aligned with the identified genes. The ones overlapping with these genes were assigned a “Brain” value of 1, the rest were assigned a “Brain” value of 0. All SNPs were subsequently assigned LD-informed “Brain” scores (see below).

Annotation of genomic regions, LD-based. The SNPs that fall within certain regions of interest may capture only a limited portion of the association signal ascribable to that region. We used an LD-weighted scoring algorithm^{35,56,71} to identify SNPs that tag specific DNA regions even if they are not situated within them. For each SNP, a pairwise correlation coefficient approximation to LD (r^2) was extracted for all 1KGP SNPs within a 1,000,000 base pairs (1 Mb). All r^2 values < 0.2 were set to 0 and each SNP was assigned an r^2 value of 1.0 with itself. LD-weighted region annotation scores for all DNA regions of interest were computed as the sum of LD r^2 between the tag SNP and all 1KGP SNPs in those regions. Given SNP i , its LD-weighted region annotation score was computed as $LDscore_i = \sum_j (\delta_j r_{ij}^2)$, where r_{ij}^2 is the LD r^2 between SNP i and SNP j and δ_j takes values of 1 or 0 depending on whether the 1KGP SNP j is within the region of interest or not. LD scores were assigned to exons, introns, 3'UTR and 5'UTR^{35,56,71,73}.

Intergenic correction. Intergenic SNPs are defined as having LD-weighted annotation scores for exon, intron, 3'UTR and 5'UTR equal to zero and being in LD with no SNPs in the 1KGP reference panel located within 100,000 base pairs of a protein coding gene, within a non-coding RNA, within a transcription factor binding site or within a miRNA binding site⁷¹. Those singled out in this way are expected to form a collection of non-genic SNPs not belonging to any annotated functional elements and their LD-associated regions within the genome and therefore represent a collection of likely null associations. Intergenic SNPs were used to estimate the inflation of GWAS summary statistics due to cryptic relatedness. We used intergenic SNPs because their relative depletion of associations suggests they provide a set of reliably null SNPs that is less contaminated by polygenic effects. The inflation factor, λ_{GC} , was estimated as the median squared z-score of independent sets of intergenic SNPs across one hundred LD-pruning iterations, divided by the expected median of a chi-square distribution with one degree of freedom.

Conditional quantile-quantile plots. To visualize enrichment, we constructed conditional quantile-quantile (Q-Q) plots where we compared the nominal p-value distribution to the empirical distribution⁷¹. In the presence of null relationships, the nominal p-values form a straight line on a Q-Q plot when plotted against the empirical distribution. We plotted $-\log_{10}$ nominal p-values against $-\log_{10}$ empirical p-values for the two SNP strata subdivided by the PNSS score, as well as for all SNPs. Leftward deflections of the observed distribution from the null line reflect increased tail probabilities in the distribution of test statistics (z-scores) and consequently an over-abundance of low p-values compared to that expected under the null hypothesis⁷¹. Enrichment is present if the line corresponding to the variants of interest has a leftward deflection from the comparison stratum. To assess polygenic effects below the standard GWAS significance threshold, we focused the Q-Q plots on SNPs with nominal $-\log_{10}(p) < 7.3$ (corresponding to $p > 5 \times 10^{-8}$).

Fold enrichment plots. To visually emphasize the association enrichment, we used conditional fold enrichment plots⁷⁴. As for Q-Q plots, the covariate of interest, i.e. the PNSS score, is used to subdivide SNPs into two strata. The plots were obtained by computing the empirical cumulative distribution of $-\log_{10}(p)$ -values for SNP association with a given phenotype for all SNPs, and for the two SNPs strata determined by the PNSS score. Then each stratum's fold enrichment was calculated as the ratio $CDF_{stratum}/CDF_{all}$ between the $-\log_{10}(p)$ cumulative distribution for that stratum and the $-\log_{10}(p)$ cumulative distribution for all SNPs. The nominal $-\log_{10}(p)$ values are plotted on the x-axis, the fold enrichment in the y-axis. To assess polygenic effects below the standard GWAS significance threshold, we focused the fold enrichment plots on SNPs with nominal $-\log_{10}(p) < 7.3$ (corresponding to $p > 5 \times 10^{-8}$). Enrichment is present if the line corresponding to the SNPs of interest has an upward deflection. The plots should be interpreted with caution when the baseline is determined by fewer than 5–10 data points.

Stratified enrichment analysis. To quantify the contribution of variants within the PNSS regions we conducted analyses using an approach⁴⁵ based on stratified LD score regression⁴⁴. We first dichotomized the PNSS scores into binary scores. We used the LD-score tool with the “-h2” option to estimate SNP-based heritability of variants with negative NSS score, controlling for a set of 53 annotations⁴⁴, including standard genomic annotations such as exon, intron, 3'UTR, 5'UTR, presence of enhancers, total LD score, and brain gene affiliation. Given the complex LD⁷⁵ in the extended major histocompatibility complex (MHC) region (genome build 19 location 25119106–33854733), we excluded SNPs in the MHC region and SNPs in LD ($r^2 > 0.1$) with such SNPs from the analysis, to avoid any inflation due to complex correlations. We then used the LD-score tool with the “-l2”

option to calculate the total LD of 1,190,321 variants from the HapMap3 project towards the category of variants with negative scores. The pairwise LD r^2 measures were calculated across 9,997,231 variants from the reference panel (1 kG Phase3 genotypes for individuals with European descent). The effect reported (β_{PNSS}) is the LD score regression coefficient. Its p-value is the probability of the “true” effect size being different from zero based on the standard error estimate of the coefficient.

Data can be obtained from. EduYears: <https://www.thessgac.org/data>, Colledge: <http://www.ccace.ed.ac.uk/node/335>, GCA1: https://ctg.cncr.nl/software/summary_statistics, GCA 2: data can be obtained from the CHARGE consortium, BMI and Height: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

References

- Heyes, C. New thinking: the evolution of human cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 2091–2096, <https://doi.org/10.1098/rstb.2012.0111> (2012).
- Whiten, A. & Erdal, D. The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 2119 (2012).
- Gunz, P., Neubauer, S., Maureille, B. & Hublin, J.-J. Brain development after birth differs between Neanderthals and modern humans. *Current Biology* **20**, R921–R922, <https://doi.org/10.1016/j.cub.2010.10.018> (2010).
- Pearce, E., Stringer, C. & Dunbar, R. I. M. New insights into differences in brain organization between Neanderthals and anatomically modern humans. *Proceedings of the Royal Society B: Biological Sciences* **280** (2013).
- d’Errico, F. *et al.* Archaeological evidence for the emergence of language, symbolism, and music - An alternative multidisciplinary perspective. *Journal of World Prehistory* **17**, 1–70, <https://doi.org/10.1023/A:1023980201043> (2003).
- Barkley, R. A. The executive functions and self-regulation an evolutionary neuropsychological perspective. *Neuropsychology Review* **11**, 1–29, <https://doi.org/10.1023/a:1009085417776> (2001).
- Wynn, T. & Coolidge, F. L. The implications of the working memory model for the evolution of modern cognition. *Int J Evol Biol* **2011**, 741357, <https://doi.org/10.4061/2011/741357> (2011).
- Ambrose, S. H. Paleolithic Technology and Human Evolution. *Science* **291**, 1748–1753, <https://doi.org/10.1126/science.1059487> (2001).
- Tomasello, M. & Herrmann, E. Ape and Human Cognition. *Current Directions in Psychological Science* **19**, 3–8, <https://doi.org/10.1177/0963721409359300> (2010).
- Finlay, B. L., Darlington, R. B. & Nicastro, N. Developmental structure in brain evolution. *Behavioral and Brain Sciences* **24**, 263–278, <https://doi.org/10.1017/s0140525x01003958> (2001).
- Dunbar, R. I. M. The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology* **32**, 163–181, <https://doi.org/10.1146/annurev.anthro.32.061002.093158> (2003).
- Pinker, S. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences* **107**, 8993–8999, <https://doi.org/10.1073/pnas.0914630107> (2010).
- Majkić, A., Evans, S., Stepanchuk, V., Tsveliykh, A. & d’Errico, F. A decorated raven bone from the Zaskalnaya VI (Kolosovskaya) Neanderthal site, Crimea. *PLoS One* **12**, e0173435, <https://doi.org/10.1371/journal.pone.0173435> (2017).
- Gintis, H. Gene–culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 878 (2011).
- van Schaik, C. P. & Burkart, J. M. Social learning and evolution: the cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **366**, 1008–1016, <https://doi.org/10.1098/rstb.2010.0304> (2011).
- Hagenaars, S. P. *et al.* Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112,151) and 24 GWAS consortia. *Mol Psychiatry* **21**, 1624–1632, <https://doi.org/10.1038/mp.2015.225> (2016).
- Sirin, S. R. Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research* **75**, 417–453, <https://doi.org/10.3102/00346543075003417> (2005).
- Pecora, P. J. *et al.* Assessing the educational achievements of adults who were formerly placed in family foster care. *Child & Family Social Work* **11**, 220–231, <https://doi.org/10.1111/j.1365-2206.2006.00429.x> (2006).
- Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471, <https://doi.org/10.1126/science.1235488> (2013).
- Baker, L. A., Treloar, S. A., Reynolds, C. A., Heath, A. C. & Martin, N. G. Genetics of educational attainment in Australian twins: Sex differences and secular changes. *Behavior Genetics* **26**, 89–102, <https://doi.org/10.1007/bf02359887> (1996).
- Davies, G. *et al.* Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* **16**, 996–1005, <https://doi.org/10.1038/mp.2011.85> (2011).
- Deary, I. J., Strand, S., Smith, P. & Fernandes, C. Intelligence and educational achievement. *Intelligence* **35**, 13–21, <https://doi.org/10.1016/j.intell.2006.02.001> (2007).
- Calvin, C. M. *et al.* Multivariate Genetic Analyses of Cognition and Academic Achievement from Two Population Samples of 174,000 and 166,000 School Children. *Behavior Genetics* **42**, 699–710, <https://doi.org/10.1007/s10519-012-9549-7> (2012).
- Gale, C. R. *et al.* Cognitive ability in early adulthood and risk of 5 specific psychiatric disorders in middle age: the Vietnam experience study. *Archives of General Psychiatry* **65**, 1410–1418, <https://doi.org/10.1001/archpsyc.65.12.1410> (2008).
- Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241, <https://doi.org/10.1038/ng.3406> (2015).
- Hill, W. D., Davies, G., Liwald, D. C., McIntosh, A. M. & Deary, I. J. Age-Dependent Pleiotropy Between General Cognitive Function and Major Psychiatric Disorders. *Biol Psychiatry* **80**, 266–273, <https://doi.org/10.1016/j.biopsych.2015.08.033> (2016).
- Smeland, O. B. *et al.* Identification of genetic loci jointly influencing schizophrenia risk and the cognitive traits of verbal-numerical reasoning, reaction time, and general cognitive function. *JAMA Psychiatry*, <https://doi.org/10.1001/jamapsychiatry.2017.1986> (2017).
- Hublin, J. J. The origin of Neandertals. *Proceedings of the National Academy of Sciences* **106**, 16022–16027, <https://doi.org/10.1073/pnas.0904119106> (2009).
- Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722, <https://doi.org/10.1126/science.1188021> (2010).
- Meyer, M. *et al.* Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507, <https://doi.org/10.1038/nature17405> (2016).
- Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**, 1103–1108, <https://doi.org/10.1038/nature04789> (2006).
- Prüfer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531, <https://doi.org/10.1038/nature11128> (2012).

33. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87, <https://doi.org/10.1038/nature04072> (2005).
34. Hill, W. D. *et al.* Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Transl Psychiatry* **6**, e980, <https://doi.org/10.1038/tp.2016.246> (2016).
35. Srinivasan, S. *et al.* Genetic Markers of Human Evolution Are Enriched in Schizophrenia. *Biological Psychiatry*, <https://doi.org/10.1016/j.biopsych.2015.10.009> (2016).
36. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49, <https://doi.org/10.1038/nature12886> (2014).
37. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N = 112 151). *Mol Psychiatry* **21**, 758–767, <https://doi.org/10.1038/mp.2016.45> (2016).
38. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542, <https://doi.org/10.1038/nature17671> (2016).
39. Davies, G. *et al.* Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N = 53,949). *Mol Psychiatry* **20**, 183–192, <https://doi.org/10.1038/mp.2014.188> (2015).
40. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, <https://doi.org/10.1038/s41588-018-0152-6> (2018).
41. Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M. & Gottesman, I. I. Just one g: consistent results from three test batteries. *Intelligence* **32**, 95–107, [https://doi.org/10.1016/s0160-2896\(03\)00062-x](https://doi.org/10.1016/s0160-2896(03)00062-x) (2004).
42. Ree, M. J. & Earles, J. A. The Stability of G across Different Methods of Estimation. *Intelligence* **15**, 271–278, [https://doi.org/10.1016/0160-2896\(91\)90036-D](https://doi.org/10.1016/0160-2896(91)90036-D) (1991).
43. Donald, M. Précis of Origins of the modern mind: Three stages in the evolution of culture and cognition. *Behavioral and Brain Sciences* **16**, 737, <https://doi.org/10.1017/s0140525x00032647> (2010).
44. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228–1235, <https://doi.org/10.1038/ng.3404> (2015).
45. Zuber, V. *et al.* Identification of shared genetic variants between schizophrenia and lung cancer. *Scientific Reports* **8**, 674, <https://doi.org/10.1038/s41598-017-16481-4> (2018).
46. Turchin, M. C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet* **44**, 1015–1019, <https://doi.org/10.1038/ng.2368> (2012).
47. Sanjak, J. S., Sidorenko, J., Robinson, M. R., Thornton, K. R. & Visscher, P. M. Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National Academy of Sciences* **115**, 151–156, <https://doi.org/10.1073/pnas.1707227114> (2018).
48. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet*, <https://doi.org/10.1038/ng.3869> (2017).
49. Marioni, R. E. *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* **44**, 26–32, <https://doi.org/10.1016/j.intell.2014.02.006> (2014).
50. Kong, A. *et al.* Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, <https://doi.org/10.1073/pnas.1612113114> (2017).
51. Subiaul, F. What's Special about Human Imitation? A Comparison with Enculturated Apes. *Behavioral Sciences* **6**, 13 (2016).
52. Herrmann, E. & Tomasello, M. Apes' and children's understanding of cooperative and competitive motives in a communicative situation. *Dev Sci* **9**, 518–529, <https://doi.org/10.1111/j.1467-7687.2006.00519.x> (2006).
53. Want, S. C. & Harris, P. L. Learning from other people's mistakes: causal understanding in learning to use a tool. *Child Development* **72**, 431–443 (2001).
54. Horner, V., Whiten, A., Flynn, E. & de Waal, F. B. Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences* **103**, 13878–13883, <https://doi.org/10.1073/pnas.0606015103> (2006).
55. Hare, B. & Tomasello, M. Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour* **68**, 571–581 (2004).
56. Srinivasan, S. *et al.* Probing the Association between Early Evolutionary Markers and Schizophrenia. *PLoS One*, <https://doi.org/10.1371/journal.pone.0169227> (2017).
57. Fu, W. & Akey, J. M. Selection and adaptation in the human genome. *Annu Rev Genomics Hum Genet* **14**, 467–489, <https://doi.org/10.1146/annurev-genom-091212-153509> (2013).
58. Burger, R. & Gimelfarb, A. Genetic variation maintained in multilocus models of additive quantitative traits under stabilizing selection. *Genetics* **152**, 807–820 (1999).
59. Pritchard, J. K. & Di Rienzo, A. Adaptation - not by sweeps alone. *Nat Rev Genet* **11**, 665–667, <https://doi.org/10.1038/nrg2880> (2010).
60. Andreassen, O. A. *et al.* Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate. *PLoS Genetics* **9**, <https://doi.org/10.1371/journal.pgen.1003455> (2013).
61. Andreassen, O. A., Thompson, W. K. & Dale, A. M. Boosting the power of schizophrenia genetics by leveraging new statistical tools. *Schizophrenia Bulletin* **40**, 13–17, <https://doi.org/10.1093/schbul/sbt168> (2014).
62. Andreassen, O. A. *et al.* Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *American Journal of Human Genetics* **92**, 197–209, <https://doi.org/10.1016/j.ajhg.2013.01.001> (2013).
63. Andreassen, O. A. *et al.* Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: implications for immune related disease mechanisms. *Mol Psychiatry* (2014).
64. Holland, D. *et al.* Estimating phenotypic polygenicity and causal effect size variance from GWAS summary statistics while accounting for inflation due to cryptic relatedness. *BioRxiv*, <https://doi.org/10.1101/133132> (2017).
65. Shadrin, A. A. *et al.* Novel Loci Associated With Attention-Deficit/Hyperactivity Disorder Are Revealed by Leveraging Polygenic Overlap With Educational Attainment. *Journal of the American Academy of Child and Adolescent Psychiatry*, <https://doi.org/10.1016/j.jaac.2017.11.013> (2018).
66. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**(8), 1112–1121, <https://doi.org/10.1038/s41588-018-0147-3> (2018).
67. Mackenbach, J. P. *et al.* Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine* **358**, 2468–2481, <https://doi.org/10.1056/NEJMsa0707519> (2008).
68. Keri, S. Genes for psychosis and creativity: a promoter polymorphism of the neuregulin 1 gene is related to creativity in people with high intellectual achievement. *Psychological Science* **20**, 1070–1073, <https://doi.org/10.1111/j.1467-9280.2009.02398.x> (2009).
69. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206, <https://doi.org/10.1038/nature14177> (2015).
70. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838, <https://doi.org/10.1038/nature09410> (2010).
71. Schork, A. J. *et al.* All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genetics* **9**, e1003449–e1003449, <https://doi.org/10.1371/journal.pgen.1003449> (2013).

72. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. *Nature Reviews Neuroscience* **11**, 201–211, <https://doi.org/10.1038/nrn2793> (2010).
73. Wang, Y. *et al.* Leveraging Genomic Annotations and Pleiotropic Enrichment for Improved Replication Rates in Schizophrenia GWAS. *PLoS Genetics* **12**, e1005803, <https://doi.org/10.1371/journal.pgen.1005803> (2016).
74. Consortium, T. E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640, <https://doi.org/10.1126/science.1105136> (2004).
75. Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *American Journal of Human Genetics* **83**, 132–135, <https://doi.org/10.1016/j.ajhg.2008.06.005> (2008).

Acknowledgements

This work was supported by the Research Council of Norway (#223273, #225989, #248778) South-East Norway Health Authority (#2016-064) and KG Jebsen Stiftelsen (#SKGJ-Med-008). GD and IJD are supported by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology funded by the Biotechnology and Biological Sciences Research Council and Medical Research Council (MR/K026992/1). WDH is supported by Age UK (Disconnected Mind grant).

Author Contributions

O.B.S. and O.A.A. designed the study. S.S. and O.F. analysed the data. S.S., F.B., O.B.S. and O.A.A. wrote the manuscript. G.D., I.J.D., W.D.H., S.D., T.U., I.M. and O.A.A. provided data. Y.W., A.W., A.J.S., R.S.D., W.K.T. and A.M.D. provided analytical tools or support. All authors commented on and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30387-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018