nature genetics

ARTICLES

https://doi.org/10.1038/s41588-020-00754-2

Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction

Perline A. Demange[1,2,3,20], Margherita Malanchini[4,5,6,20], Travis T. Mallard[6], Pietro Biroli[7], Simon R. Cox[8], Andrew D. Grotzinger[6], Elliot M. Tucker-Drob[6,9], Abdel Abdellaoui[1,10], Louise Arseneault[5], Elsje van Bergen[1,3], Dorret I. Boomsma[1], Avshalom Caspi[5,11,12,13], David L. Corcoran[12], Benjamin W. Domingue[14], Kathleen Mullan Harris[15], Hill F. Ip[1], Colter Mitchell[16], Terrie E. Moffitt[5,11,12,13], Richie Poulton[17], Joseph A. Prinz[12], Karen Sugden[11], Jasmin Wertz[11], Benjamin S. Williams[11], Eveline L. de Zeeuw[1,3], Daniel W. Belsky[18,19,21] ✉, K. Paige Harden[6,21] ✉ and Michel G. Nivard[1,21] ✉

Little is known about the genetic architecture of traits affecting educational attainment other than cognitive ability. We used genomic structural equation modeling and prior genome-wide association studies (GWASs) of educational attainment ($n = 1{,}131{,}881$) and cognitive test performance ($n = 257{,}841$) to estimate SNP associations with educational attainment variation that is independent of cognitive ability. We identified 157 genome-wide-significant loci and a polygenic architecture accounting for 57% of genetic variance in educational attainment. Noncognitive genetics were enriched in the same brain tissues and cell types as cognitive performance, but showed different associations with gray-matter brain volumes. Noncognitive genetics were further distinguished by associations with personality traits, less risky behavior and increased risk for certain psychiatric disorders. For socioeconomic success and longevity, noncognitive and cognitive-performance genetics demonstrated associations of similar magnitude. By conducting a GWAS of a phenotype that was not directly measured, we offer a view of genetic architecture of noncognitive skills influencing educational success.

It takes something more than intelligence to act intelligently.

Fyodor Dostoyevsky, *Crime and Punishment*

Success in school—and life—depends on skills beyond cognitive ability[1–4]. Randomized trials of early life educational interventions find substantial benefits to educational outcomes, employment and adult health, even though the interventions have no lasting effects on children's cognitive functions[5,6]. These results have captured the attention of educators and policy-makers, motivating interest in so-called 'noncognitive skills'[7–9]. Noncognitive skills suspected to be important for educational success include motivation, curiosity, persistence and self-control[1,10–13]. However, questions have been raised about the substance of these skills and the magnitudes of their impacts on life outcomes[14].

Twin studies find evidence that noncognitive skills are heritable[3,15–18]. Genetic analysis could help clarify the contribution of these skills to educational attainment and elucidate their connections with other traits. However, lack of consistent and reliable measurements of noncognitive skills in existing genetic datasets poses challenges[19].

To overcome these challenges, we designed a GWAS of a latent trait, that is, a trait not measured in any of the genotyped subjects[20]. We borrowed the strategy used in the original analysis of noncognitive skills within the discipline of economics[21,22]: we defined genetic influences on noncognitive skills as the genetic variation in educational attainment that was not explained by cognitive skills. We then performed GWASs on this residual 'noncognitive' genetic variation in educational attainment. This approach is a necessarily imperfect representation of the true relationship between cognitive and noncognitive skills; in human development, cognitive abilities and other skills relevant for educational attainment probably interact dynamically, each influencing the other[23]. Our analysis excludes genetic

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2]Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, the Netherlands. [3]Research Institute LEARN!, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [4]Department of Biological and Experimental Psychology, Queen Mary University of London, London, UK. [5]Social, Genetic and Developmental Psychiatric Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [6]Department of Psychology, University of Texas at Austin, Austin, TX, USA. [7]Department of Economics, University of Zurich, Zurich, Switzerland. [8]Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, Edinburgh, UK. [9]Population Research Center, University of Texas at Austin, Austin, TX, USA. [10]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands. [11]Department of Psychology & Neuroscience, Duke University, Durham, NC, USA. [12]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA. [13]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [14]Stanford Graduate School of Education, Stanford University, Palo Alto, CA, USA. [15]Department of Sociology and Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [16]Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [17]Department of Psychology and Dunedin Multidisciplinary Health and Development Research Unit, University of Otago, Dunedin, New Zealand. [18]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA. [19]Robert N. Butler Columbia Aging Center, Columbia University, New York, NY, USA. [20]These authors contributed equally: Perline A. Demange, Margherita Malanchini. [21]These authors jointly supervised this work: Daniel W. Belsky, K. Paige Harden, Michel G. Nivard. ✉e-mail: daniel.belsky@columbia.edu; harden@utexas.edu; m.g.nivard@vu.nl

influences on education-relevant skills that also influence measured cognitive abilities. The value of this imperfect approach is to make a quantity otherwise difficult to study tractable for analysis.

We conducted analysis using genomic structural equation modeling (Genomic-SEM)[24] applied to published GWAS summary statistics for educational attainment and cognitive performance[25]. Our analysis used these summary statistics to 'subtract' genetic influence on cognitive performance from the association of each SNP with educational attainment. The remaining associations of each SNP with educational attainment formed a new GWAS of a noncognitive skills phenotype that was never directly measured. We call this new statistical approach GWAS-by-subtraction.

We used results from the GWAS-by-subtraction of noncognitive skills to conduct two sets of analyses. First, we conducted hypothesis-driven analysis using the phenotypic annotation approach[26]. We used genetic correlation and polygenic score analysis to test the hypothesis that noncognitive skills influence educational and economic attainments and longevity and to investigate traits and behaviors that constitute noncognitive skills. Second, we conducted hypothesis-free bioinformatic annotation analysis to explore the tissues, cell types and brain structures that might distinguish the biology of noncognitive skills from the biology mediating cognitive influences on educational attainment.

Results

GWAS-by-subtraction identifies genetic associations with noncognitive variance in educational attainment. The term 'noncognitive skills' was originally coined by economists studying individuals who were equivalent in cognitive ability but differed in educational attainment[22]. Our analysis of noncognitive skills was designed to mirror this original approach: we focused on genetic variation in educational outcomes not explained by genetic variation in cognitive ability. Specifically, we applied Genomic-SEM[24] to summary statistics from GWASs of educational attainment[25] and cognitive performance[25]. Both phenotypes were regressed on a latent factor representing genetic variance in cognitive performance (hereafter '*Cog*'). Educational attainment was further regressed on a second latent factor representing the residual genetic variance in educational attainment left over after regressing out variance related to cognitive performance (hereafter '*NonCog*'). By construction, *NonCog* genetic variance was independent of *Cog* genetic variance ($r_g = 0$). In other words, the *NonCog* factor represents genetic variation in educational attainment that is not accounted for by the *Cog* factor. These two latent factors were then regressed on individual SNPs, yielding a GWAS of the latent constructs *NonCog* and *Cog*. A graphic representation of the model is presented in Fig. 1. Parameters are derived in terms of the observed moments of the joint distribution of educational attainment, cognitive performance and an SNP (Supplementary Note).

The *NonCog* latent factor accounted for 57% of total genetic variance in educational attainment. Using linkage disequilibrium (LD) score regression[27], we estimated the SNP heritability for *NonCog* to be $h^2_{NonCog} = 0.0637$ (s.e. = 0.0021). After conventional GWAS significance threshold correction, GWAS of *NonCog* identified 157 independent genome-wide-significant lead SNPs (independent SNPs defined as outside a 250-kb window, or within a 250-kb window and LD $r^2 < 0.1$). The results from the *NonCog* GWAS are shown as a Manhattan plot in Fig. 2. *NonCog* and *Cog* GWAS details are reported in Supplementary Tables 1–4, Supplementary Fig. 1 and the Supplementary Note. In addition, we report a series of sensitivity analyses as follows: analysis of potential biases due to cohort differences (Supplementary Table 5 and Supplementary Figs. 2–4); analysis of impact of allowing for positive genetic correlations between *NonCog* and *Cog* (Supplementary Tables 6 and 7, and Supplementary Figs. 5 and 6); and analysis of impact of allowing for a moderate causal effect of educational attainment on cognitive performance[28] (Supplementary Table 8 and Supplementary Figs. 7–9).

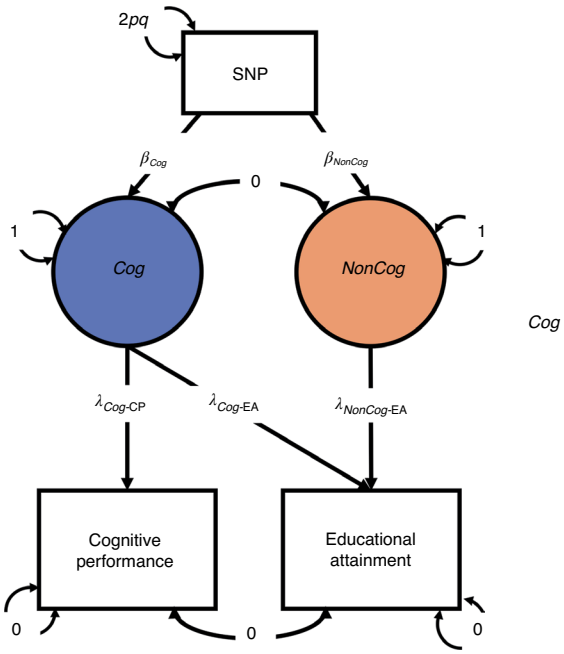

Fig. 1 | GWAS-by-subtraction Genomic-SEM model. Cholesky model as fitted in Genomic-SEM, with path estimates for a single SNP included as illustration. SNP, cognitive performance (CP) and educational attainment (EA) are observed variables based on GWAS summary statistics. The genetic covariance between CP and EA is estimated based on their GWAS summary statistics. The model is fitted to a 3 × 3 observed variance–covariance matrix (that is, SNP, CP, EA). *Cog* and *NonCog* are latent (unobserved) variables. The covariances between CP and EA and between *Cog* and *NonCog* are fixed to 0. The variance of the SNP is fixed to the value of 2*pq* (*p* = reference allele frequency, *q* = alternative allele frequency, based on 1000 Genomes Project phase 3). The residual variances of CP and EA are fixed to 0, so that all variance is explained by the latent factors. The variances of the latent factors are fixed to 1. The observed variables CP and EA were regressed on the latent variables, resulting in the estimates for the path loadings: $\lambda_{Cog\text{-}CP} = 0.4465$; $\lambda_{Cog\text{-}EA} = 0.2237$; $\lambda_{NonCog\text{-}EA} = 0.2565$. The latent variables were then regressed on each SNP that met quality control criteria.

Phenotypic annotation analysis elucidates correlates of noncognitive skills genetics. Our phenotypic annotation analyses proceeded in two steps. First, we conducted polygenic score (PGS) and genetic correlation (rG) analysis to test whether our GWAS-by-subtraction succeeded in identifying genetic influences that were important to educational attainment and also distinct from genetic influences on cognitive ability. Second, we conducted PGS and rG analyses to explore how *NonCog* related to a network of phenotypes that psychology and economics research suggests might form the basis of noncognitive influences on educational attainment.

***NonCog* genetics are associated with education, socioeconomic attainment and longevity**. To establish whether the Genomic-SEM GWAS-by-subtraction succeeded in isolating genetic variance in education that was independent of cognitive function, we compared genetic associations of *NonCog* and *Cog* with educational attainment and cognitive test performance. Results for analysis of education and cognitive test phenotypes are shown in Fig. 3.

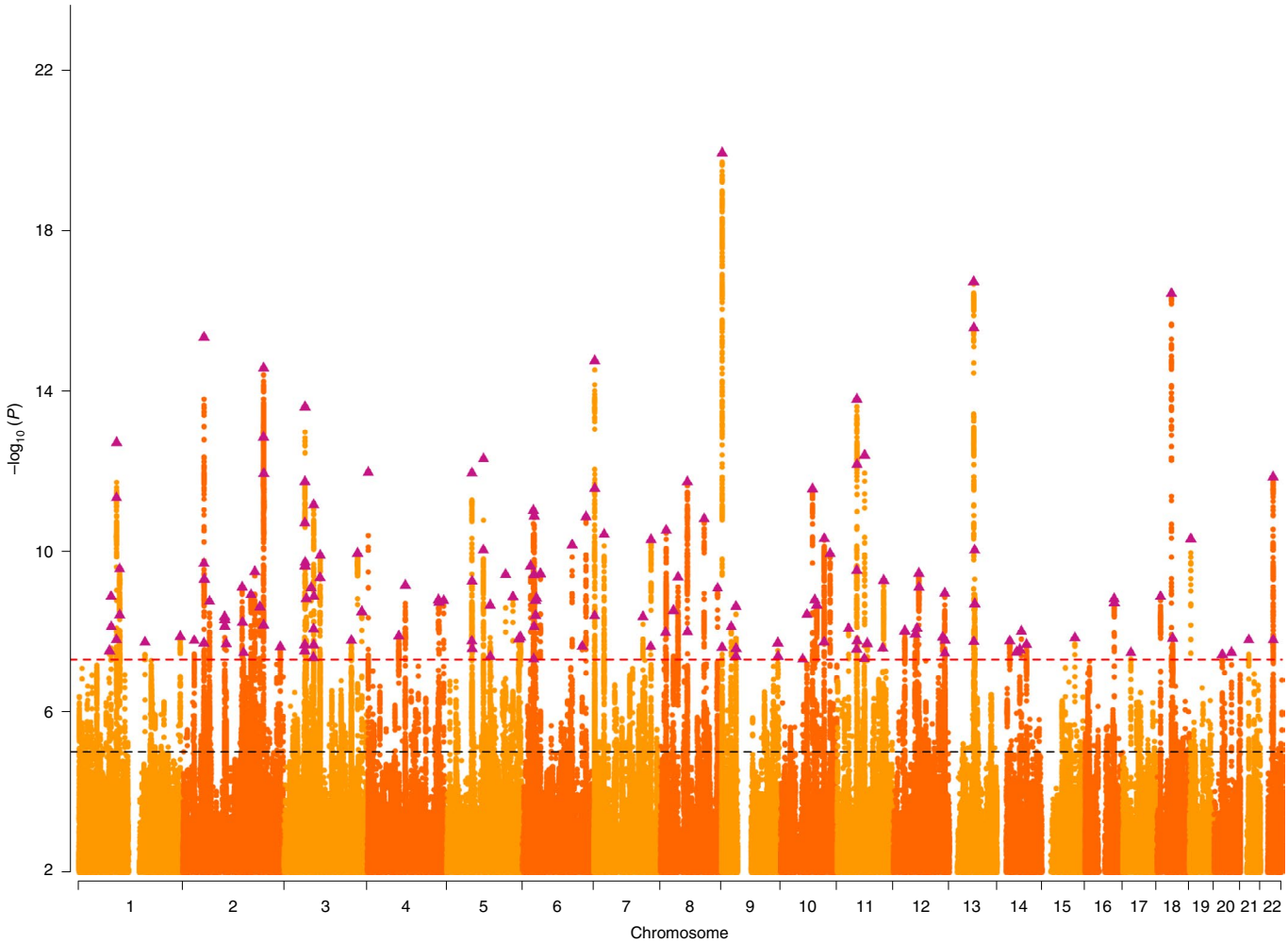

Fig. 2 | Manhattan plot of SNP associations with *NonCog*. Plot of the $-\log_{10}$(*P* value) associated with Wald's test (two-sided) of β_{NonCog} for all SNPs, ordered by chromosome and base position. Purple triangles indicate genome-wide-significant ($P<5\times10^{-8}$) and independent (within a 250-kb window and $r^2<0.1$) associations. The red dashed line marks the threshold for genome-wide significance ($P=5\times10^{-8}$) and the black dashed line the threshold for nominal significance ($P=1\times10^{-5}$).

We conducted PGS analysis of educational attainment in the Netherlands Twin Register[29] (NTR), National Longitudinal Study of Adolescent to Adult Health[30] (AddHealth), Dunedin Longitudinal Study[31], E-Risk[32] and Wisconsin Longitudinal Study[33] (WLS) cohorts (meta-analysis $n=24{,}056$; cohort descriptions in Supplementary Tables 9 and 10 and Supplementary Note). PGS effect sizes were the same for *NonCog* and *Cog* (*NonCog* $\beta=0.24$ (s.e.$=0.03$), *Cog* $\beta=0.24$ (s.e.$=0.02$), $P_{diff}=0.702$; all PGS results are reported in Supplementary Tables 11 and 12). We conducted complementary genetic correlation analysis using Genomic-SEM and GWAS summary statistics from a hold-out sample GWAS of educational attainment (Supplementary Note). This analysis allowed us to compute an out-of-sample genetic correlation of *NonCog* with educational attainment. *NonCog* showed a stronger genetic correlation with educational attainment compared with *Cog* (*NonCog* $r_g=0.71$ (s.e.$=0.02$), *Cog* $r_g=0.57$ (s.e.$=0.02$), $P_{diff}<0.0001$; all genetic correlation results are reported in Supplementary Tables 13 and 14).

We conducted PGS analysis of cognitive test performance in the NTR, Texas Twin Project[34], Dunedin, E-Risk and WLS cohorts (combined $n=11{,}351$). The goal of our GWAS-by-subtraction analysis was to exclude, as much as possible, genetic variance in cognitive ability from genetic variance in skills relevant for education. Consistent with this goal, effect sizes for *NonCog* PGS associations with full-scale intelligence quotient (IQ) were smaller by half compared with *Cog* PGS associations (*NonCog* $\beta=0.17$ (s.e.$=0.02$), *Cog* $\beta=0.29$ (s.e.$=0.03$); $P_{diff}<0.0001$). However, the non-zero correlation between the *NonCog* PGS and full-scale IQ is a reminder that the cognitive performance GWAS used in our GWAS-by-subtraction analyses does not capture the entirety of genetic influences on all forms of cognitive tests measured at all points in the lifespan. Additional PGS analyses of IQ subscales are reported in Supplementary Fig. 10 and Supplementary Tables 11 and 12.

We conducted complementary genetic correlation analysis using results from a published GWAS of childhood IQ[35]. Parallel to PGS analysis, the *NonCog* genetic correlation with childhood IQ was smaller by more than half compared with the *Cog* genetic correlation (*NonCog* $r_g=0.31$ (s.e.$=0.06$), *Cog* $r_g=0.75$ (s.e.$=0.08$), $P_{diff\_fdr}<0.0001$). Of the total genetic correlation between childhood IQ and educational attainment, 31% of the covariance was explained by *NonCog* and 69% by *Cog*.

We next examined downstream economic and health outcomes associated with greater educational attainment[36,37]. In the PGS analysis in the AddHealth and Dunedin cohorts ($n=6{,}358$), *NonCog* and *Cog* PGSs showed similar associations with occupational attainment (*NonCog* $\beta=0.21$ (s.e.$=0.01$), *Cog* $\beta=0.21$ (s.e.$=0.01$), $P_{diff}=0.902$). In genetic correlation analysis, *NonCog* showed a

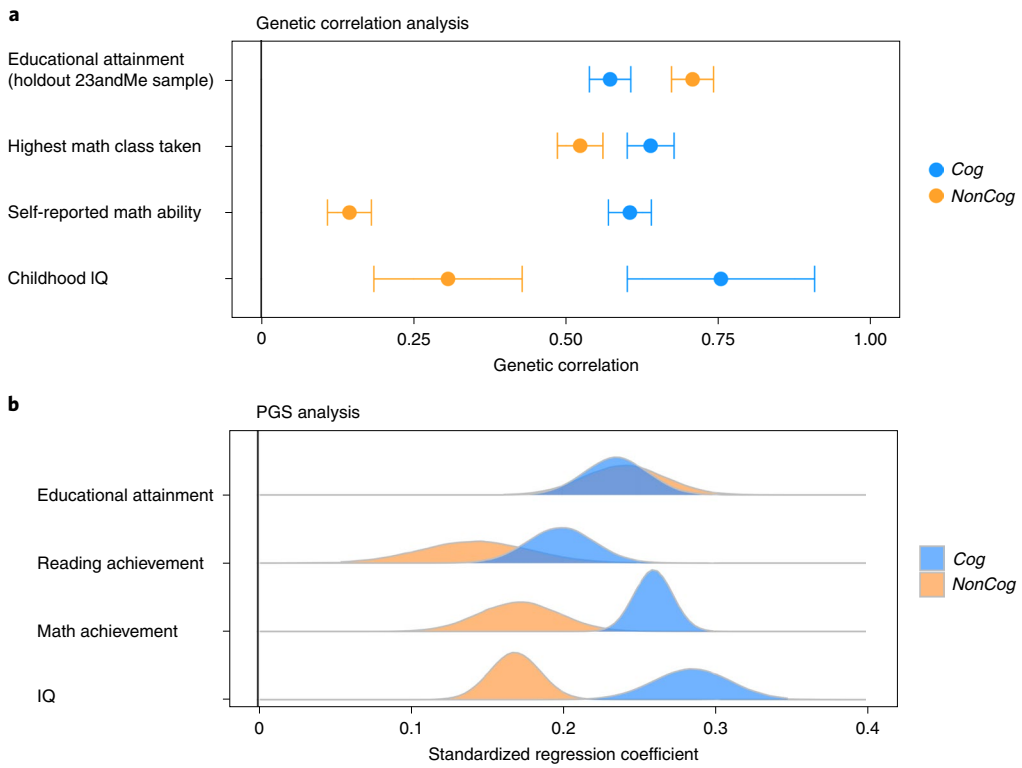

Fig. 3 | Polygenic prediction and genetic correlations with IQ and educational achievement. a, Genetic correlations of *NonCog* and *Cog* with educational attainment, highest math class taken, self-reported math ability and childhood IQ. The dots represent genetic correlations estimated using Genomic-SEM. Correlations with *NonCog* are in orange and with *Cog* in blue. Error bars represent 95% confidence intervals (CIs). Exact estimates and *P* values are reported in Supplementary Table 14. For analysis of genetic correlations with educational attainment, we re-ran the Genomic-SEM model to compute *NonCog* and *Cog* using summary statistics that omitted the 23andMe sample from the educational attainment GWAS. We then used the 23andMe sample to run the GWAS of educational attainment. Thus, there is no sample overlap in this analysis. **b**, Effect-size distributions from meta-analysis of *NonCog* and *Cog* PGS associations with cognitive test performance and educational attainment. Outcomes were regressed simultaneously on *NonCog* and *Cog* PGSs. Effect sizes entered into the meta-analysis were standardized regression coefficients interpretable as Pearson's *r*. Exact estimates and *P* values are reported in Supplementary Table 12. Samples and measures are detailed in Supplementary Tables 9 and 10. Traits were measured in different samples: educational attainment was measured in the AddHealth, Dunedin, E-Risk, NTR and WLS samples ($n = 24{,}056$); reading achievement and mathematics achievement were measured in the AddHealth, NTR and Texas Twin samples ($n = 9{,}274$ for reading achievement; $n = 10{,}747$ for mathematics achievement); cognitive test performance (IQ) was measured in the Dunedin, E-Risk, NTR, Texas Twins and WLS samples ($n = 11{,}351$). The densities were obtained by randomly generating normal distributions where the meta-analytic estimate was included as the mean and the meta-analytic s.e. as the s.d.

similar relationship to income[38] to *Cog* (*NonCog* $r_g = 0.62$ (s.e. = 0.04), *Cog* $r_g = 0.62$ (s.e. = 0.04), $P_{diff\_fdr} = 0.947$) and a stronger relationship with neighborhood deprivation[38], a measure related to where a person can afford to live (*NonCog* $r_g = -0.51$ (s.e. = 0.05), *Cog* $r_g = -0.32$ (s.e. = 0.04), $P_{diff\_fdr} = 0.001$). In Genomic-SEM analysis, *NonCog* explained 53% of the genetic correlation between educational attainment and income, and 65% of the genetic correlation between educational attainment and neighborhood deprivation (Supplementary Table 15).

We conducted genetic correlation analysis of longevity based on GWASs of parental lifespan[39]. Genetic correlations were stronger for *NonCog* compared with *Cog* (*NonCog* $r_g = 0.37$ (s.e. = 0.03); *Cog* $r_g = 0.27$ (s.e. = 0.03); $P_{diff\_fdr} = 0.024$). In Genomic-SEM analysis, *NonCog* explained 61% of the genetic correlation between educational attainment and longevity.

In summary, *NonCog* and *Cog* genetics showed similar relationships with educational attainment and its long-term outcomes, despite *NonCog* genetics having a much weaker relationship to measured cognitive test performance than *Cog* genetics. These findings broadly support the hypothesis that noncognitive skills, distinct from cognitive abilities, are an important contributor to success across the life course.

We next conducted a series of genetic correlation analyses to explore the network of phenotypes to which *NonCog* was genetically correlated. To develop understanding of the substance of noncognitive skills, we tested where in that network of phenotypes genetic correlations with *NonCog* diverged from genetic correlations with *Cog*. Our analysis was organized around four themes: decision-making preferences, health-risk and fertility behaviors, personality traits and psychiatric disorders. The results of genetic correlation analyses are shown in Fig. 4 and Supplementary Fig. 11, and reported in Supplementary Table 14.

***NonCog* genetics were associated with decision-making preferences**. In economics, noncognitive influences on achievement and health are often studied in relation to decision-making preferences[40–43]. *NonCog* was genetically correlated with higher tolerance of risks[44] ($r_g = 0.10$ (s.e. = 0.03)) and willingness to forgo immediate gratification in favor of a larger reward at a later time[45] (delay discounting $r_g = -0.52$ (s.e. = 0.08)). In contrast, *Cog* was genetically correlated with generally more cautious decision-making characterized by lower levels of risk tolerance ($r_g = -0.35$ (s.e. = 0.07), $P_{diff\_fdr} < 0.0001$) and delayed discounting ($r_g = -0.35$ (s.e. = 0.07), $P_{diff\_fdr} = 0.082$).

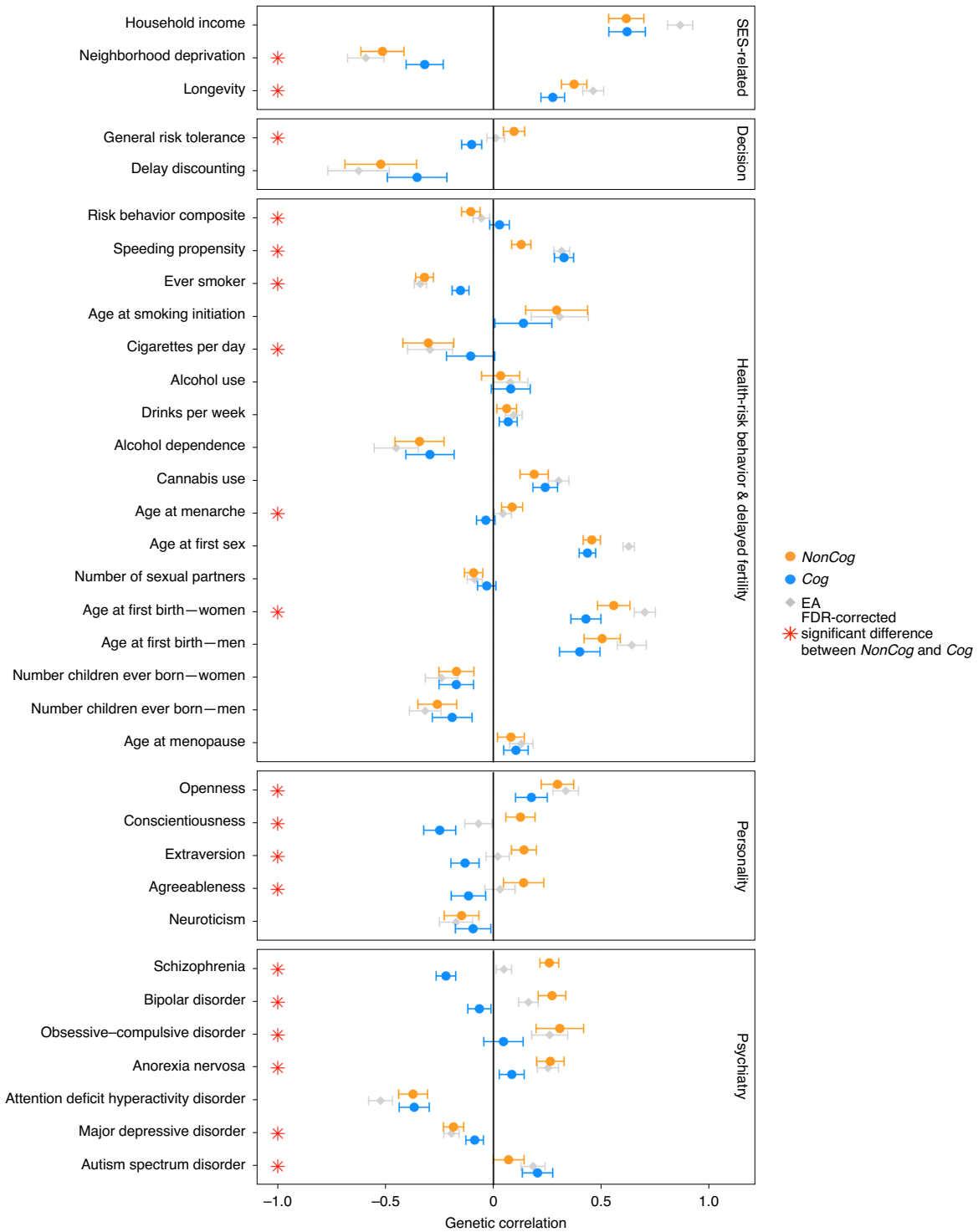

Fig. 4 | Estimates of genetic correlations with *NonCog*, *Cog* and educational attainment. Genetic correlations of *NonCog*, *Cog* and educational attainment with selected phenotypes. The dots represent genetic correlations estimated in Genomic-SEM. Correlations with *NonCog* are in orange, with *Cog* in blue and with educational attainment in gray. Error bars represent 95% CIs. Red stars indicate a statistically significant (FDR-corrected $P < 0.05$, two-tailed test) difference in the magnitude of the correlation with *NonCog* versus *Cog*. Exact *P* values for all associations are reported in Supplementary Table 14. The FDR correction was applied based on all genetic correlations tested (including in Supplementary Fig. 11). The difference test is based on a χ^2 test associated with a comparison between a model constraining these two correlations to be identical versus a model where the correlations are freely estimated. SES, socioeconomic status. Source GWASs are listed in Supplementary Table 13.

***NonCog* genetics were associated with less health-risk behavior and delayed fertility.** An alternative approach to studying specific noncognitive skills is to infer individual differences in noncognitive skills from patterns of health-risk behavior. *NonCog* was genetically correlated with less health-risk behavior, as indicated by analysis of obesity[46], substance use[44,47–50], and sexual behaviors and early

fertility[44,51,52] (r_g range 0.2–0.5), with the exception that the r_g with alcohol use was not different from zero and r_g with cannabis use was positive. Genetic correlations for *Cog* were generally in the same direction but of smaller magnitude.

***NonCog* genetics were associated with personality characteristics linked with social and professional competency**. In psychology, noncognitive influences on achievement are conceptualized as personality traits, that is, patterns of stable individual differences in emotion and behavior. The model of personality that has received the most attention in genetics is a five-factor model referred to as the Big Five. Genetic correlation analysis of the Big Five personality traits[53–55] revealed that *NonCog* genetics were most strongly associated with Openness to Experience (being curious and eager to learn; $r_g = 0.30$ (s.e. = 0.04)) and were further associated with a pattern of personality characteristic of changes that occur as people mature in adulthood[56]. Specifically, *NonCog* showed a positive r_g with Conscientiousness (being industrious and orderly; $r_g = 0.13$ (s.e. = 0.03)), Extraversion (being enthusiastic and assertive; $r_g = 0.14$ (s.e. = 0.03)), and Agreeableness (being polite and compassionate; $r_g = 0.14$ (s.e. = 0.05)), and negative r_g with Neuroticism (being emotionally volatile; $r_g = -0.15$ (s.e. = 0.04)). Genetic correlations of *Cog* with Openness to Experience and Neuroticism were similar to those for *NonCog* ($P_{diff\_fdr\text{-}Openness} = 0.040$, $P_{diff\_fdr\text{-}Neuroticism} = 0.470$). In contrast, genetic correlations of *Cog* with Conscientiousness, Extraversion and Agreeableness were in the opposite direction ($r_g = -0.25$ to -0.12, $P_{diff\_fdr} < 0.0005$). PGS analysis of personality traits is reported in Supplementary Table 12, Supplementary Fig. 12 and the Supplementary Note.

***NonCog* genetics were associated with higher risk for multiple psychiatric disorders**. In clinical psychology and psychiatry, research is focused on mental disorders. Mental disorders are generally associated with impairments in academic achievement and social role functioning[57,58]. However, positive genetic correlations with educational attainment and creativity have been reported for some disorders[59,60]. We therefore tested *NonCog* r_g with psychiatric disorders based on published case–control GWASs of mental disorders[61–67]. *NonCog* was associated with higher risk for multiple clinically defined disorders, including anorexia nervosa ($r_g = 0.26$ (s.e. = 0.04)), obsessive–compulsive disorder ($r_g = 0.31$ (s.e. = 0.06)), bipolar disorder ($r_g = 0.27$ (s.e. = 0.03)) and schizophrenia ($r_g = 0.26$ (s.e. = 0.02)). Genetic correlations between *Cog* and psychiatric disorders were either smaller in magnitude (anorexia nervosa $r_g = 0.08$ (s.e. = 0.03), $P_{diff\_fdr} < 0.001$; obsessive–compulsive disorder $r_g = 0.05$ (s.e. = 0.05), $P_{diff\_fdr} = 0.002$) or in the opposite direction (bipolar disorder $r_g = -0.07$ (s.e. = 0.03), $P_{diff\_fdr} < 0.001$; schizophrenia $r_g = -0.22$ (s.e. = 0.02), $P_{diff\_fdr} < 0.001$). Both *NonCog* and *Cog* showed negative genetic correlations with attention deficit hyperactivity disorder (*NonCog* $r_g = -0.37$ (s.e. = 0.03), *Cog* $r_g = -0.37$ (s.e. = 0.04), $P_{diff\_fdr} = 0.947$).

In summary, *NonCog* genetics were associated with phenotypes from economics and psychology thought to mediate noncognitive influences on educational success. These associations contrasted with associations for *Cog* genetics, supporting distinct pathways of influence on achievement in school and later in life. Opposing patterns of association were also observed for psychiatric disorders, suggesting that the unexpected positive genetic correlation between educational attainment and mental health problems uncovered in previous studies[60,68,69] arises from noncognitive genetic influences on educational attainment.

Biological annotation analyses reveal shared and specific neurobiological correlates. The goal of biological annotation of GWAS discoveries is to elucidate molecular mechanisms mediating genetic influences on the phenotype of interest. Our biological annotation analysis proceeded in two steps. First, we conducted enrichment analysis to test whether some tissues and cell types were more likely to mediate *NonCog* and *Cog* heritabilities than others. Second, we conducted genetic correlation analysis to explore how *NonCog* and *Cog* genetics related to different brain structures.

***NonCog* and *Cog* genetics were enriched in similar tissues and cells**. We tested whether common variants in genes specifically expressed in 53 Genotype-Tissue Expression (GTEx) tissues[70] or in 152 tissues captured in a previous aggregation of RNA-sequencing studies[71,72] were enriched in their effects on *Cog* or *NonCog*. Genes predominantly expressed in the brain rather than peripheral tissues were enriched in both *NonCog* and *Cog* (Supplementary Table 16).

To examine expression patterns at a more granular level of analysis, we used MAGMA[73] and stratified the LD score regression (LDSC)[74] to test enrichment of common variants in 265 nervous system, cell-type-specific gene sets[75] (Supplementary Table 17). In MAGMA analysis, common variants in 95 of 265 gene sets were enriched for association with *NonCog*. The enriched cell types were predominantly neurons (97%), with enrichment most pronounced for telencephalon-projecting neurons, di- and mesencephalon neurons, and, to a lesser extent, telencephalon interneurons (Supplementary Fig. 13 and Supplementary Table 18). Enrichment for *Cog* was similar to *NonCog* (correlation between *z*-statistics Pearson's $r = 0.85$), and there were no differences in cell-type-specific enrichment, suggesting that the same types of brain cells mediate genetic influences on *NonCog* and *Cog* (Supplementary Fig. 14). Stratified LDSC results were similar to results from MAGMA (Supplementary Note, Supplementary Fig. 15 and Supplementary Table 19).

The absence of differences in cell-type-specific enrichment is surprising given that *NonCog* and *Cog* are genetically uncorrelated. We therefore used the TWAS/Fusion tool[76] to conduct gene-level analysis. This analysis revealed a mixture of concordant and discordant gene effects on *NonCog* and *Cog* consistent with the genetic correlation of 0 (Supplementary Note, Supplementary Fig. 16 and Supplementary Table 20).

***NonCog* and *Cog* genetics show diverging associations with total and regional brain volumes**. Educational attainment has previously been found to be genetically correlated with greater total brain volume[77,78]. We therefore used a GWAS of regional brain volume to compare the r_g of *NonCog* and *Cog* with total brain volume and 100 regional brain volumes (99 gray-matter volumes and 1 white-matter volume) controlling for total brain volume (Supplementary Table 21)[79]. For total brain volume, genetic correlation was stronger for *Cog* compared with *NonCog* (*Cog* $r_g = 0.22$ (s.e. = 0.04), *NonCog* $r_g = 0.07$ (s.e. = 0.03), $P_{diff} = 0.005$). Total gray-matter volume, controlling for total brain volume, was not associated with either *NonCog* or *Cog* (*NonCog*: $r_g = 0.07$ (s.e. = 0.04); *Cog*: $r_g = 0.06$ (s.e. = 0.04)). For total white-matter volume, conditional on total brain volume, genetic correlation was weakly negative for *NonCog* compared with *Cog* (*NonCog* $r_g = -0.12$ (s.e. = 0.04), *Cog* $r_g = -0.01$ (s.e. = 0.04), $P_{diff} = 0.04$).

NonCog was not associated with any of the regional gray-matter volumes after false discovery rate (FDR) correction. In contrast, *Cog* was significantly associated with regional gray-matter volumes for the bilateral fusiform, insula and posterior cingulate (r_g range 0.11–0.17), as well as the left superior temporal ($r_g = 0.11$ (s.e. = 0.04)), left pericalcarine ($r_g = -0.16$ (s.e. = 0.05)) and right superior parietal volumes ($r_g = -0.22$ (s.e. = 0.06)) (Fig. 5).

Finally, we tested genetic correlation of *NonCog* and *Cog* with white-matter tract integrity as measured using diffusion tensor imaging (DTI)[80]. Analyses included 5 DTI parameters in each of 22 white-matter tracts (Supplementary Table 22). *NonCog* was positively associated with the mode of anisotropy (MO) parameter (which denotes a more tubular, as opposed to planar, water

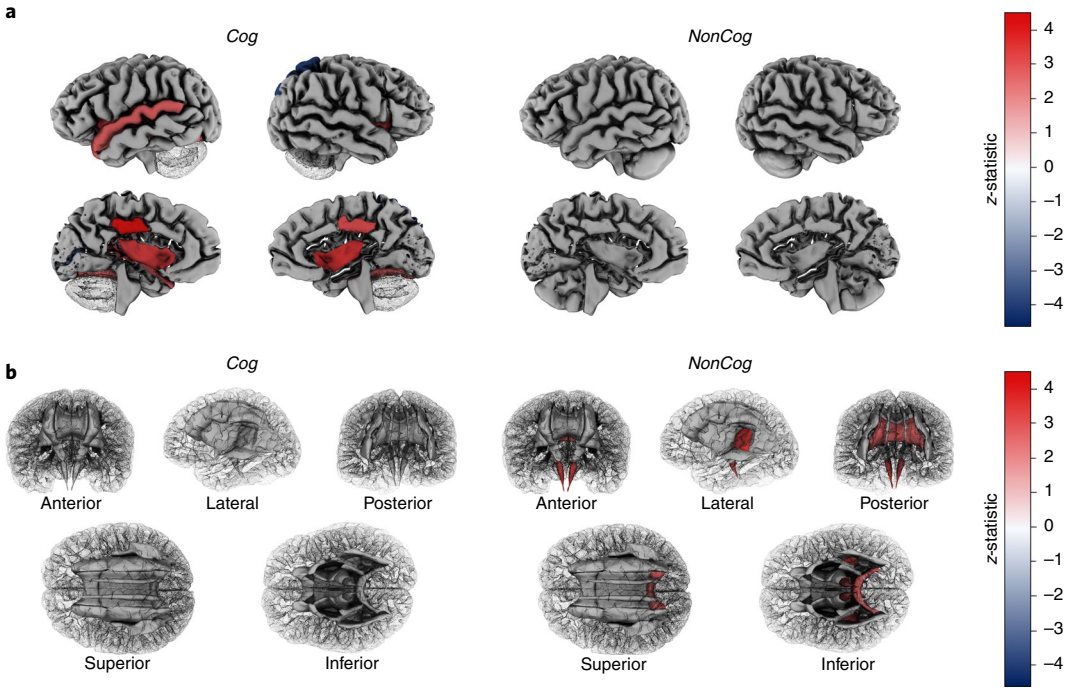

Fig. 5 | Genetic correlations with regional gray-matter volumes and white-matter tracts. a, Cortical patterning of FDR-corrected significant genetic correlations with regional gray-matter volumes for *Cog* versus *NonCog*, after correction for total brain volume. Regions of interest are plotted according to the Desikan–Killiany–Tourville atlas[97], shown on a single, manually edited surface (http://mindboggle.info)[98]. Exact estimates and *P* values are reported in Supplementary Table 21. *Cog* showed significant associations with gray-matter volume for the bilateral fusiform, insula and posterior cingulate, the left superior temporal, and left pericalcarine and right superior parietal volumes. *NonCog* was not associated with any of the regional brain volumes. **b**, White-matter tract patterning of FDR-corrected. significant genetic correlations with regional MO for *Cog* versus *NonCog*. White-matter tract probability maps are plotted according to the Johns Hopkins University DTI atlas (https://identifiers.org/neurovault.image:1401)[99]. Exact estimates and *P* values are reported in Supplementary Table 21. *Cog* was not associated with regional MO. *NonCog* showed significant associations with MO in the corticospinal tract, the retrolenticular limb of the internal capsule and the splenium of the corpus callosum.

diffusion) in the corticospinal tract, retrolenticular limb of the internal capsule and splenium of the corpus callosum (Fig. 5). However, all correlations were small ($0.10 < r_g < 0.14$), and we detected no genetic correlations that differed between *NonCog* and *Cog* (Supplementary Note).

Discussion

GWASs of noncognitive influences on educational attainment identified 157 independent loci and polygenic architecture, accounting for more than half the genetic variance in educational attainment. In genetic correlation and PGS analysis, these *NonCog* genetics showed a similar magnitude of associations with educational attainment, economic attainment and longevity to genetics associated with cognitive influences on educational attainment (*Cog*). As expected, *NonCog* genetics had much weaker associations with cognition phenotypes compared with *Cog* genetics. These results contribute new GWAS evidence in support of the hypothesis that heritable noncognitive skills influence educational attainment and downstream life-course economic and health outcomes.

Phenotypic and biological annotation analyses shed light on the substance of heritable noncognitive skills influencing education. Economists hypothesize that preferences that guide decision-making in the face of risk and delayed rewards represent noncognitive influences on educational attainment. Consistent with this hypothesis, *NonCog* genetics were associated with higher risk tolerance and lower time discounting. These decision-making preferences are associated with financial wealth, whereas the opposite preferences are hypothesized to contribute to a feedback loop perpetuating poverty[81]. Consistent with results from the analysis of decision-making preferences, *NonCog* genetics were also associated with healthier behavior and later fertility.

Psychologists hypothesize that the Big Five personality characteristics of conscientiousness and openness are the two 'pillars of educational success'[2,3,82]. Our results provide some support for this hypothesis, with the strongest genetic correlation evident for openness. However, they also show that noncognitive skills encompass the full range of personality traits, including agreeableness, extraversion and the absence of neuroticism. This pattern mirrors the pattern of personality change that occurs as young people mature into adulthood[56]. Thus, noncognitive skills share genetic etiology with what might be termed a 'mature personality'. The absolute magnitudes of genetic correlations between *NonCog* and individual personality traits are modest. This result suggests that the personality traits described by psychologists capture some, but not all, genetic influence on noncognitive skills.

Although the general pattern of findings in our phenotypic annotation analysis indicated that noncognitive skills were genetically related to socially desirable characteristics and behaviors, there was an important exception. Genetic correlation analysis of a psychiatric disorder GWAS revealed positive associations of *NonCog* genetics with schizophrenia, bipolar disorder, anorexia nervosa and obsessive–compulsive disorder. Previously, these psychiatric disorders have been shown to have a positive r_g with educational attainment, a result that has been characterized as paradoxical given the impairments in educational and occupational functioning typical of serious mental illness. Our results clarify that these associations are driven by noncognitive factors associated with success in education. These results align with the theory that clinically defined psychiatric

disorders represent extreme manifestations of dimensional psychological traits, which might be associated with adaptive functioning within the normal range[83–85].

Finally, biological annotation analyses suggested that genetic variants contributing to educational attainment not mediated through cognitive abilities are enriched in genes expressed in the brain, specifically in neurons. Even though *NonCog* and *Cog* were genetically uncorrelated, variants in the same neuron-specific gene sets were enriched for both traits. Although we found some evidence of differences between *NonCog* and *Cog* in associations with gray-matter volumes, moderate sample sizes in neuroimaging GWASs mean that these results must be treated as preliminary, requiring replication with data from larger-scale GWASs of white-matter and gray-matter phenotypes. Limited differentiation of *NonCog* and *Cog* in biological annotation analyses focused at the levels of tissue and cell type highlights the need for finer-grained molecular data resources to inform these analyses and the complementary value of phenotypic annotation analyses focused at the level of psychology and behavior.

We acknowledge limitations. Cognitive and noncognitive skills develop in interaction with each other. For example, the dynamic mutualism hypothesis[86] proposes that noncognitive characteristics shape investments of time and effort, leading to differences in the pace of cognitive development[87,88]. However, in Genomic-SEM analysis, the *NonCog* factor is, by construction, uncorrelated with genetic influences on adult cognition as measured in the *Cog* GWAS. Our statistical separation of *NonCog* from cognition is thus a simplified representation of development. Longitudinal studies with repeated measures of cognitive and candidate noncognitive skills are needed to study their reciprocal relationships across development[89,90]. Our statistical separation of *NonCog* from cognition is also incomplete. The ability to control statistically for any variable, genetic or otherwise, depends on how well and comprehensively that variable is measured[91]. The tests of cognitive performance included in the *Cog* GWAS probably do not capture all genetic influences on all forms of cognitive ability across the lifespan[92,93]. Despite these limitations, our simplified and incomplete statistical separation of *NonCog* from *Cog* allowed us to test whether heritable traits other than cognitive ability influenced educational attainment and to explore what those traits might be.

As our analysis was based on GWAS of educational attainment, noncognitive genetics identified in the present study may differ from noncognitive genetics affecting other socioeconomic attainments such as income, or traits and behaviors that mediate responses to early childhood interventions, to the extent that those genetics do not affect educational attainment. Parallel analysis of alternative attainment phenotypes will clarify the specificity of discovered noncognitive genetics.

In the case of GWASs of educational attainment, the included samples were drawn mainly from western Europe and the USA, and participants completed their education in the late twentieth and early twenty-first centuries. The phenotype of educational attainment reflects an interaction between an individual and the social system in which they are educated. Differences across social systems, including education policy, culture and historical context, may result in different heritable traits influencing educational attainment[94]. Results therefore may not generalize beyond the times and places where GWAS samples were collected.

Generalization of the *NonCog* factor is also limited by restriction of the included GWASs to individuals of European ancestry. Lack of methods for integrating genome-scale genetic data across populations with different ancestries[95,96] requires this restriction, but raises threats to external validity. GWASs of other ancestries and development of methods for trans-ancestry analysis can enable analysis of (*Non)Cog* in non-European populations.

Within the bounds of these limitations, the results illustrate the application of Genomic-SEM to conduct GWASs of a phenotype not directly measured in GWAS databases. This application could have broad utility beyond the genetics of educational attainment. The GWAS-by-subtraction method allowed us to study a previously hard-to-interpret residual value. Our analysis provides a view of the genetic architecture of noncognitive skills influencing educational success. These skills are central to theories of human capital formation within the social and behavioral sciences and are increasingly the targets of social policy interventions. Our results establish that noncognitive skills are central to the heritability of educational attainment and illuminate connections between genetic influences on these skills and social and behavioral science phenotypes.

Online content

Any methods, additional references, Nature Research reporting summaries, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-00754-2.

Received: 19 March 2020; Accepted: 19 November 2020;
Published online: 7 January 2021

References

1. Moffitt, T. E. et al. A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl Acad. Sci. USA* **108**, 2693–2698 (2011).
2. von Stumm, S., Hell, B. & Chamorro-Premuzic, T. The hungry mind: intellectual curiosity is the third pillar of academic performance. *Perspect. Psychol. Sci.* **6**, 574–588 (2011).
3. Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D. & Harden, K. P. Genetically-mediated associations between measures of childhood character and academic achievement. *J. Pers. Soc. Psychol.* **111**, 790–815 (2016).
4. Heckman, J. J., Stixrud, J. & Urzua, S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *J. Labor Econ.* **24**, 411–482 (2006).
5. Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. & Yavitz, A. The rate of return to the HighScope Perry Preschool Program. *J. Public Econ.* **94**, 114–128 (2010).
6. Conti, G., Heckman, J. J. & Pinto, R. The effects of two influential early childhood interventions on health and healthy behaviour. *Econ. J.* **126**, F28–F65 (2016).
7. Gutman, L. M. & Schoon, I. The impact of non-cognitive skills on outcomes for young people. *Educ. Endow. Found.* **59**, 2019 (2013).
8. Garcia, E. *The Need to Address Noncognitive Skills in the Education Policy Agenda* (Economic Policy Institute, 2014).
9. Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B. & Borghans, L. Fostering and *Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success* OECD Education Working Papers No. 110 (OECD Publishing, 2014).
10. Heckman, J. J. Skill formation and the economics of investing in disadvantaged children. *Science* **312**, 1900–1902 (2006).
11. Heckman, J. J. & Kautz, T. Hard evidence on soft skills. *Labour Econ.* **19**, 451–464 (2012).
12. Rimfeld, K., Kovas, Y., Dale, P. S. & Plomin, R. True grit and genetics: predicting academic achievement from personality. *J. Pers. Soc. Psychol.* **111**, 780–789 (2016).
13. Richardson, M., Abraham, C. & Bond, R. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* **138**, 353–387 (2012).
14. Smithers, L. G. et al. A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat. Hum. Behav.* **2**, 867–880 (2018).
15. Kovas, Y. et al. Why children differ in motivation to learn: insights from over 13,000 twins from 6 countries. *Personal. Individ. Differ.* **80**, 51–63 (2015).
16. Loehlin, J. C. *Genes and Environment in Personality Development* (Sage Publications, 1992).
17. Tucker-Drob, E. M. & Harden, K. P. Learning motivation mediates gene-by-socioeconomic status interaction on mathematics achievement in early childhood. *Learn. Individ. Differ.* **22**, 37–45 (2012).
18. Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P. & Tucker-Drob, E. M. 'Same but different': associations between multiple aspects of self-regulation, cognition, and academic abilities. *J. Pers. Soc. Psychol.* **117**, 1164–1188 (2019).

19. Morris, T. T., Smith, G. D., van Den Berg, G. & Davies, N. M. Investigating the longitudinal consistency and genetic architecture of non-cognitive skills, and their relation to educational attainment. Preprint at *bioRxiv* https://doi.org/10.1101/470682 (2019).
20. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case–control association mapping by proxy using family history of disease. *Nat. Genet.* **49**, 325–331 (2017).
21. Bowles, S. & Gintis, H. *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life* (Basic Books, 1977).
22. Heckman, J. J. & Rubinstein, Y. The importance of noncognitive skills: lessons from the GED Testing Program. *Am. Econ. Rev.* **91**, 145–149 (2001).
23. Ackerman, P. L., Kanfer, R. & Goff, M. Cognitive and noncognitive determinants and consequences of complex skill acquisition. *J. Exp. Psychol. Appl.* **1**, 270–304 (1995).
24. Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
25. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
26. Belsky, D. W. & Harden, K. P. Phenotypic annotation: using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Curr. Dir. Psychol. Sci.* **28**, 82–90 (2019).
27. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
28. Ritchie, S. J. & Tucker-Drob, E. M. How much does education improve intelligence? A meta-analysis. *Psychol. Sci.* **29**, 1358–1369 (2018).
29. Ligthart, L. et al. The Netherlands Twin Register: longitudinal research based on twin and twin-family designs. *Twin Res. Hum. Genet.* **22**, 623–636 (2019).
30. Harris, K. M. et al. Cohort profile: the National Longitudinal Study of Adolescent to Adult Health (AddHealth). *Int. J. Epidemiol.* **48**, 1415–1415k (2019).
31. Poulton, R., Moffitt, T. E. & Silva, P. A. The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. *Soc. Psychiatry Psychiatr. Epidemiol.* **50**, 679–693 (2015).
32. Moffitt, T. E., E-risk Team. Teen-aged mothers in contemporary Britain. *J. Child Psychol. Psychiatry* **43**, 727–742 (2002).
33. Herd, P., Carr, D. & Roan, C. Cohort profile: Wisconsin Longitudinal Study (WLS). *Int. J. Epidemiol.* **43**, 34–41 (2014).
34. Harden, K. P., Tucker-Drob, E. M. & Tackett, J. L. The Texas Twin Project. *Twin Res. Hum. Genet.* **16**, 385–390 (2013).
35. Benyamin, B. et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol. Psychiatry* **19**, 253–258 (2014).
36. Chetty, R. et al. The association between income and life expectancy in the United States, 2001–2014. *JAMA* **315**, 1750–1766 (2016).
37. Case, A. & Deaton, A. Mortality and morbidity in the 21st century. *Brook. Pap. Econ. Act.* **2017**, 397–476 (2017).
38. Hill, W. D. et al. Molecular genetic contributions to social deprivation and household income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).
39. Timmers, P. R. et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife* **8**, e39856 (2019).
40. Almlund, M., Duckworth, A. L., Heckman, J. & Kautz, T. in *Handbook of the Economics of Education* Vol. 4 (eds Hanushek, E. A., Machin, S. & Wößman, L.) 1–181 (Elsevier, 2011).
41. Borghans, L., Duckworth, A. L., Heckman, J. J. & Weel, Bter The economics and psychology of personality traits. *J. Hum. Resour.* **43**, 972–1059 (2008).
42. Rabin, M. A perspective on psychology and economics. *Eur. Econ. Rev.* **46**, 657–685 (2004).
43. Becker, A., Deckers, T., Dohmen, T., Falk, A. & Kosse, F. The relationship between economic preferences and psychological personality measures. *Annu. Rev. Econ.* **4**, 453–478 (2012).
44. Linnér, R. K. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
45. Sanchez-Roige, S. et al. Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nat. Neurosci.* **21**, 16–18 (2018).
46. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
47. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
48. Walters, R. K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
49. Schumann, G. et al. KLB is associated with alcohol drinking, and its gene product β-Klotho is necessary for FGF21 regulation of alcohol preference. *Proc. Natl Acad. Sci. USA* **113**, 14372–14377 (2016).
50. Pasman, J. A. et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. *Nat. Neurosci.* **21**, 1161–1170 (2018).
51. Linnér, R. K. et al. Multivariate genomic analysis of 1.5 million people identifies genes related to addiction, antisocial behavior, and health. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.16.342501 (2020).
52. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
53. Lo, M.-T. et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017).
54. John, O. P., Naumann, L. P. & Soto, C. J. in *Handbook of Personality: Theory and Research* (eds John, O. P. et al.) 114–158 (Guilford Press, 2008).
55. de Moor, M. H. M. et al. Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* **17**, 337–349 (2012).
56. Caspi, A., Roberts, B. W. & Shiner, R. L. Personality development: stability and change. *Annu. Rev. Psychol.* **56**, 453–484 (2005).
57. Kessler, R. C. et al. Social consequences of psychiatric disorders, I: educational attainment. *Am. J. Psychiatry* **152**, 1026–1032 (1995).
58. Breslau, J., Lane, M., Sampson, N. & Kessler, R. C. Mental disorders and subsequent educational attainment in a US national sample. *J. Psychiatr. Res.* **42**, 708–716 (2008).
59. Power, R. A. et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat. Neurosci.* **18**, 953–955 (2015).
60. Bansal, V. et al. Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nat. Commun.* **9**, 3078 (2018).
61. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
62. Ruderfer, D. M. et al. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018).
63. Jansen, P. R. et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
64. Duncan, L. et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174**, 850–858 (2017).
65. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
66. Arnold, P. D. et al. Revealing the complex genetic architecture of obsessive–compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
67. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
68. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
69. Nieuwboer, H. A., Pool, R., Dolan, C. V., Boomsma, D. I. & Nivard, M. G. GWIS: genome-wide inferred statistics for functions of multiple phenotypes. *Am. J. Hum. Genet.* **99**, 917–927 (2016).
70. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
71. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
72. Fehrmann, R. S. N. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
73. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
74. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
75. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
76. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
77. Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W. & Koellinger, P. D. Are bigger brains smarter? Evidence from a large-scale preregistered study. *Psychol. Sci.* **30**, 43–54 (2019).
78. Elliott, M. L. et al. A polygenic score for higher educational attainment is associated with larger brains. *Cereb. Cortex* **29**, 3496–3504 (2019).
79. Zhao, B. et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.* **51**, 1637–1644 (2019).

80. Zhao, B. et al. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol. Psychiatry* https://doi.org/10.1038/s41380-019-0569-z (2019).
81. Haushofer, J. & Fehr, E. On the psychology of poverty. *Science* **344**, 862–867 (2014).
82. Briley, D. A., Domiteaux, M. & Tucker-Drob, E. M. Achievement-relevant personality: relations with the Big Five and validation of an efficient instrument. *Learn. Individ. Differ.* **32**, 26–39 (2014).
83. Smoller, J. W. et al. Psychiatric genetics and the structure of psychopathology. *Mol. Psychiatry* **24**, 409–420 (2019).
84. Plomin, R., Haworth, C. M. A. & Davis, O. S. P. Common disorders are quantitative traits. *Nat. Rev. Genet.* **10**, 872–878 (2009).
85. Meehl, P. E. Schizotaxia, schizotypy, schizophrenia. *Am. Psychol.* **17**, 827–838 (1962).
86. von Stumm, S. & Ackerman, P. L. Investment and intellect: a review and meta-analysis. *Psychol. Bull.* **139**, 841–869 (2013).
87. Tucker-Drob, E. M. & Harden, K. P. in *Genetics, Ethics and Education* (eds Grigorenko, E. L. et al.) 134–158 (Cambridge University Press, 2017).
88. Tucker-Drob, E. M. in *Handbook of Competence and Motivation: Theory and Application* 2nd edn (eds Elliot, A., Dweck, C. & Yeager, D.) 471–486 (Guilford Press, 2017).
89. Tucker-Drob, E. M. & Harden, K. P. Intellectual interest mediates gene × socioeconomic status interaction on adolescent academic achievement: intellectual interest and G×E. *Child Dev.* **83**, 743–757 (2012).
90. Malanchini, M. et al. Reading self-perceived ability, enjoyment and achievement: a genetically informative study of their reciprocal links over time. *Dev. Psychol.* **53**, 698–712 (2017).
91. Westfall, J. & Yarkoni, T. Statistically controlling for confounding constructs is harder than you think. *PLoS ONE* **11**, e0152719 (2016).
92. de la Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M. & Deary, I. J. A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-020-00936-2 (2020).
93. Tucker-Drob, E. M. & Briley, D. A. Continuity of genetic and environmental influences on cognition across the life span: a meta-analysis of longitudinal twin and adoption studies. *Psychol. Bull.* **140**, 949–979 (2014).
94. Tropf, F. C. et al. Hidden heritability due to heterogeneity across seven populations. *Nat. Hum. Behav.* **1**, 757–765 (2017).
95. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
96. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
97. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 171 (2012).
98. Klein, A. Mindboggle-101 manually labeled individual brains. *Harvard Dataverse, V2* https://doi.org/10.7910/DVN/HMQKCK (2016).
99. Gorgolewski, K. J. et al. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinform.* **9**, 8 (2015).

Methods

Meta-analysis of educational attainment GWASs. We reproduced the Social Science Genetic Association Consortium (SSGAC) 2018 GWAS of educational attainment[25] by meta-analyzing published summary statistics for $n = 766{,}345$ (www.thessgac.org/data) with summary statistics obtained from 23andMe, Inc. ($n = 365{,}538$). We included SNPs with sample size >500,000 and minor allele frequency > 0.005 in the 1000 Genomes Project reference set (10,101,243 SNPs). We did not apply genomic control, as standard errors of publicly available and 23andMe summary statistics were already corrected[25]. Meta-analysis was performed using METAL[100].

GWAS-by-subtraction. The objective of our GWAS-by-subtraction analysis was to estimate, for each SNP, the association with educational attainment that was independent of that SNP's association with cognition (hereafter, the *NonCog* SNP effect). We used Genomic-SEM[24] in R v.3.4.3 to analyze GWAS summary statistics for the educational attainment and cognitive performance phenotypes in the SSGAC's 2018 GWAS[25]. The model regressed the educational attainment and cognitive performance summary statistics on two latent variables, *Cog* and *NonCog* (Fig. 1). *Cog* and *NonCog* were then regressed on each SNP in the genome. This analysis allowed for two paths of association with educational attainment for each SNP. One path was fully mediated by *Cog*. The other path was independent of *Cog* and measured the noncognitive SNP effect, *NonCog*. To identify independent hits with $P < 5 \times 10^{-8}$ (the customary *P*-value threshold to approximate an α value of 0.05 in GWAS), we pruned the results using a radius of 250 kb and an LD threshold of $r^2 < 0.1$ (Supplementary Tables 1–3). We explore alternative lead SNPs and definitions of loci in Supplementary Table 4. The parameters estimated in a GWAS-by-subtraction and their derivation in terms of the genetic covariance are described in the Supplementary Note (model specification), and practical analysis steps are further described in the Supplementary Note (SNP filtering). The effective sample size of the *NonCog* and *Cog* GWAS was estimated to 510,795 and 257,700, respectively (Supplementary Note). We investigated biases from unaccounted-for heterogeneity in overlap across SNPs in the educational attainment and cognitive performance GWASs, and describe a possible strategy to deal with it (Supplementary Note). We investigated potential biases due to cohort differences in SNP heritability in the Supplementary Note. We evaluated the consequences of modifying $r_g(NonCog, Cog) = 0$ by evaluating $r_g = 0.1$, 0.2 or 0.3, and we investigated the consequences of a violation of the assumed causation between cognitive performance and educational attainment in the Supplementary Note.

Genetic correlations. We used Genomic-SEM to compute genetic correlations of *Cog* and *NonCog* with other education-linked traits for which well-powered GWAS data were available (SNP-h^2 *z*-statistics > 2; Supplementary Table 13) and to test whether genetic correlations with these traits differed between *Cog* and *NonCog*. Specifically, models tested the null hypothesis that trait genetic correlations with *Cog* and *NonCog* could be constrained to be equal using a χ^2 test with FDR adjustment to correct for multiple testing. The FDR adjustment was conducted across all genetic correlation analyses reported in the article, excluding the analyses of brain volumes described below. Finally, we used Genomic-SEM analysis of genetic correlations to estimate the percentage of the genetic covariance between educational attainment and the target traits that was explained by *Cog* and *NonCog*, using the model illustrated in Supplementary Fig. 17.

PGS analysis. PGS analyses were conducted in data drawn from six population-based cohorts from the Netherlands, the UK, the USA and New Zealand: (1) the NTR[29,101], (2) E-Risk[32], (3) the Texas Twin Project[34], (4) the AddHealth[30,102], dbGaP accession no. phs001367.v1.p1; (5) WLS[33], dbGaP accession no. phs001157.v1.p1; and (6) the Dunedin Multidisciplinary Health and Development Study[31]. Supplementary Tables 9 and 10 describe cohort-specific metrics, and we include a short description of the cohorts' populations and recruitment in Supplementary Note. Only participants with European ancestry were included in the analysis, due to the low portability of PGSs between different ancestry populations. PGSs were computed with PLINK based on weights derived using the LD-pred[103] software with an infinitesimal prior and the 1000 Genomes Project phase 3 sample as a reference for the LD structure. LD-pred weights were computed in a shared pipeline to ensure comparability between cohorts. Each outcome (for example, IQ score) was regressed on the *Cog* and *NonCog* PGSs and a set of control variables (sex, 10 principal components derived from the genetic data, and, for cohorts in which these quantities varied, genotyping chip and age), using Stata 14 for WLS, Stata 15 for E-Risk and the Dunedin study, and R (v.3.4.3 and newer) for NTR, AddHealth and the Texas Twin Project. In cohorts containing related individuals, nonindependence of observations from relatives was accounted for using generalized estimation equations or by clustering of standard errors at the family level. We used a random effects meta-analysis to aggregate the results across the cohorts. This analysis allows a cohort-specific random intercept. Individual cohort results are in Supplementary Table 11 and meta-analytic estimates in Supplementary Table 12.

Biological annotation. *Enrichment of tissue-specific gene expression.* We used gene sets defined in Finucane et al.[104] to test for the enrichment of genes specifically expressed in 1 of 53 GTEx tissues[70], or 152 tissues captured by the Franke et al. aggregation of RNA-sequencing studies[71,72]. This analysis seeks to confirm the role of brain tissues in mediating *Cog* and *NonCog* influences on educational attainment. The exact analysis pipeline used is available online (https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses).

Enrichment of cell-type-specific expression. We leveraged single-cell RNA-sequencing data of cells sampled from the mouse nervous system[75] to identify cell-type-specific RNA expression. Zeisel et al.[75] sequenced cells obtained from 19 regions in the contiguous anatomical regions in the peripheral sensory, enteric and sympathetic nervous systems. After initial quality control, they retained 492,949 cells, which were sampled down to 160,796 high-quality cells. These cells were further grouped into clusters representing 265 broad cell types. We analyzed the dataset published by Zeisel et al.[75] containing mean transcript counts for all genes with count >1 for each of the 265 clusters (Supplementary Table 17). We restricted analysis to genes with expression levels above the 25th percentile. For each gene in each cell type, we computed the cell-type-specific proportion of reads for the gene (normalizing the expression within cell type). We then computed the proportion of proportions over the 265 cell types (computing the specificity of the gene to a specific cell type). We ranked the 12,119 genes retained in terms of specificity to each cell type and then retained the 10% of genes most specific to a cell type as the 'cell-type-specific' gene set. We then tested whether any of the 265 cell-type-specific gene sets were enriched in the *Cog* or *NonCog* GWASs. This analysis sought to identify specific cell types and specific regions in the brain involved in the etiology of *Cog* and *NonCog*. We further computed the difference in enrichment for *Cog* and *NonCog* to test whether any cell types were specific to either trait. For these analyses, we leveraged two widely used enrichment analysis tools: MAGMA[73] and stratified LDSC[74] with the European reference panel from the 1000 Genomes Project phase 3 as SNP location and LD structure reference, Gencode release 19 as gene location reference and the human–mouse homology reference from MGI (http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt).

MAGMA. We used MAGMA (v.1.07b)[73], a program for gene-set analysis based on GWAS summary statistics. We computed gene-level association statistics using a window of 10 kb around the gene for both *Cog* and *NonCog*. We then used MAGMA to run a competitive gene-set analysis, using the gene *P* values and gene-correlation matrix (reflecting LD structure) produced in the gene-level analysis. The competitive gene-set analysis tests whether the genes within the cell-type-specific gene set described in Enrichment of cell-type-specific expression are more strongly associated with *Cog*/*NonCog* than other genes.

Stratified LDSC. We used LDSC to compute LD scores for the SNPs in each of our 'cell-type-specific' gene sets. Parallel to MAGMA analysis, we added a 10-kb window around each gene. We ran partitioned LDSC to compute the contribution of each gene set to the heritability of *Cog* and *NonCog*. To guard against inflation, we used LD score best practices, and included the LD score baseline model (baselineLD.v2.2) in the analysis. We judged the statistical significance of the enrichment based on the *P* value associated with the tau coefficient.

Difference in enrichment between Cog and NonCog. To compute differences in enrichment, we compute a standardized difference between the per-annotation enrichment for *Cog* and *NonCog* as:

$$Z_{\text{diff}} = \frac{e_{Cog} - e_{NonCog}}{\text{sqrt}\left(\text{s.e.}^2_{Cog} + \text{s.e.}^2_{NonCog} - 2 \times \text{CTI} \times \text{s.e.}_{Cog} \times \text{s.e.}_{NonCog}\right)} \quad (1)$$

where e_{Cog} is the enrichment of a particular gene set for *Cog*, e_{NonCog} is the enrichment for the same gene set for *NonCog*, s.e._{Cog} is the standard error of the enrichment for *Cog*, s.e._{NonCog} is the standard error of the enrichment for *NonCog* and CTI is the LD score cross-trait intercept, a metric of dependence between the GWASs of *Cog* and *NonCog*.

We investigated the significance of the difference between *Cog* and *NonCog* tau coefficient with equation (1) as well as by computing jack-knifed standard errors. From the jack-knifed estimates of the coefficient output by the LDSC software, we computed the jack-knifed estimates and standard errors of the difference between *Cog* and *NonCog* tau coefficients, as well as a *z*-statistic for each annotation.

Enrichment of gene expression in the brain. We performed a transcriptome-wide association study (TWAS) using FUSION[76] (http://gusevlab.org/projects/fusion). We used pre-computed brain–gene-expression weights available on the FUSION website, generated from 452 human individuals as part of the CommonMind Consortium. We then superimposed the bivariate distribution of the results of the TWAS for *Cog* and *NonCog* over the bivariate distribution expected, given the sample overlap between educational attainment and cognitive performance (the GWAS on which our GWASs of *Cog* and *NonCog* are based; Supplementary Note).

Brain modalities. *Brain volumes.* We conducted genetic correlation analysis of brain volumes using GWAS results published by Zhao et al.[79], who performed

GWASs of total brain volume and 100 regional brain volumes, including 99 gray-matter volumes and total white-matter volume (Supplementary Table 21). Analyses included covariate adjustment for sex, age, their square interaction and 20 principal components. Analyses of regional brain volumes additionally included covariate adjustment for total brain volume. GWAS summary statistics for these 101 brain volumes were obtained from https://med.sites.unc.edu/bigs2/data/gwas-summary-statistics. Summary statistics were filtered and pre-processed using Genomic-SEM's 'munge' function, retaining all HapMap3 SNPs with allele frequency >0.01 outside the major histocompatibility complex region. We used Genomic-SEM to compute the genetic correlations of *Cog, NonCog* and brain volumes. Analyses of regional volumes controlled for total brain volume. For each volume, we tested whether correlations differed between *Cog* and *NonCog*. Specifically, we used a χ^2 test to evaluate the null hypothesis that the two genetic correlations were equal. We used FDR adjustment to correct for multiple testing. The FDR adjustment is applied to the results for all gray-matter volumes for *Cog* and *NonCog* separately.

White-matter structures. We conducted genetic correlation analysis of white-matter structures using GWAS results published by Zhao et al.[80], who performed GWASs of DTI measures of the integrity of white-matter tracts. DTI parameters were derived for fractional anisotropy, mean diffusivity, axial diffusivity, radial diffusivity and MO. Each of these parameters was measured for 22 white-matter tracts of interests (Supplementary Table 22), resulting in 110 GWASs. GWAS summary statistics for these 110 GWASs were obtained from https://med.sites.unc.edu/bigs2/data/gwas-summary-statistics. Summary statistics were filtered and processed using Genomic-SEM's 'munge' function, retaining all HapMap3 SNPs with allele frequency >0.01 outside the major histocompatibility complex region. For each white-matter structure, we tested whether genetic correlations differed between *Cog* and *NonCog*. Specifically, we used a χ^2 test to evaluate the null hypothesis that the two genetic correlations were equal. We used FDR adjustment to correct for multiple testing. As these different diffusion parameters are statistically and logically interdependent, having been derived from the same tensor, FDR adjustment was applied to the results for each type of white-matter diffusion parameter separately. FDR correction was applied separately for *Cog* and *NonCog*.

Additional resources. A FAQ on why, how and what we studied is available at https://medium.com/@kph3k/investigating- the-genetic-architecture-of-non-cognitive- skills-using-gwas-by-subtraction-b8743773ce44. A tutorial on how to perform GWAS-by-subtraction is available at http://rpubs.com/MichelNivard/565885

Additional resources to Genomic-SEM software include:

- A wiki including numerous tutorials: https://github.com/MichelNivard/GenomicSEM/wiki
- A Genomic-SEM user group for specific questions relating to models and software: https://groups.google.com/g/genomic-sem-users
- A venue to report technical issues: https://github.com/MichelNivard/GenomicSEM/issues

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

GWAS summary data for *NonCog* and *Cog* (excluding 23andMe) have been deposited in the GWAS catalog with accession nos. GCST90011874 and GCST90011875, respectively (*NonCog* GWAS: ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011874, *Cog* GWAS: ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011875). For 23andMe dataset access, see https://research.23andme.com/dataset-access. Part of the AddHealth data is publicly available and can be downloaded at the following link: https://data.cpc.unc.edu/projects/2/view#public_li. For restricted access data, details of the data-sharing agreement and data access requirements can be found at the following link: https://data.cpc.unc.edu/projects/2/view. The Dunedin study datasets reported in the current article are not publicly available due to lack of informed consent and ethical approval, but are available on request by qualified scientists. Requests require a concept paper describing the purpose of data access, ethical approval at the applicant's university and provision for secure data access. We offer secure access on the Duke, Otago and King's College campuses. All data analysis scripts and results files are available for review (https://moffittcaspi.trinity.duke.edu/research-topics/dunedin). The E-Risk Longitudinal Twin Study datasets reported in the current article are not publicly available due to lack of informed consent and ethical approval, but are available on request by qualified scientists. Requests require a concept paper describing the purpose of data access, ethical approval at the applicant's university and provision for secure data access. We offer secure access on the Duke and King's College campuses. All data analysis scripts and results files are available for review (https://moffittcaspi.trinity.duke.edu/research-topics/erisk). NTR data may be accessed, on approval of the data access committee (email: ntr.datamanagement.fgb@vu.nl). Researchers will be able to obtain Texas Twins data through managed access. Requests for managed access should be sent to E. Tucker-Drob (tuckerdrob@utexas.edu) and P. Harden (harden@utexas.edu), joint principal investigators of the Texas Twin Project. The WLS data can be requested following this form: https://www.ssc.wisc.edu/wlsresearch/data/Request_Genetic_Data_28_June_2017.pdf.

Code availability

Code used to run the analyses is available at https://github.com/PerlineDemange/non-cognitive.

A tutorial on how to perform GWAS-by-subtraction is available at http://rpubs.com/MichelNivard/565885. All additional software used to perform these analyses is available online.

References

100. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
101. Willemsen, G. et al. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Res. Hum. Genet.* **16**, 271–281 (2013).
102. Highland, H. M., Avery, C. L., Duan, Q., Li, Y. & Harris, K. M. *Quality Control Analysis of AddHealth GWAS Data* https://www.cpc.unc.edu/projects/addhealth/documentation/guides/AH_GWAS_QC.pdf (2018).
103. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
104. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

Acknowledgements

The present study was developed with support from the Jacobs Foundation at a meeting organized by D.W.B. and K.P.H.; E.M.T.-D. and C.M. provided support, and co-authors B.W.D., J.W., M.G.N. and P.B. also attended. We thank K. Mannik and F. Tropf for contributions to the meeting, and the Jacobs Foundation Fellowship team, who made the meeting possible. D.W.B., K.P.H., M.G.N., E.M.T.-D. and C.M. are fellows of the Foundation. J.W. is a Jacobs Foundation Young Scholar. The present study used GWAS summary statistics published by the SSGAC and additional data obtained from 23andMe. We thank the research participants and employees of 23andMe for making this work possible. We thank the SSGAC and COGENT consortia for sharing their summary statistics of the GWASs of educational attainment and cognitive performance, especially A. Okbay for her quick and repeated help with providing these data. The present study used data from the NTR, the Texas Twin Study, the AddHealth, the Dunedin Longitudinal Study, the E-Risk Study and the WLS. The NTR is supported by: 'Twin-family database for behavior genetics and genomics studies' (NWO 480-04-004), Longitudinal data collection from teachers of Dutch twins and their siblings (NWO-481-08-011); Twin-family study of individual differences in school achievement (NWO 056-32-010) and Gravitation program of the Dutch Ministry of Education, Culture and Science and the Netherlands Organization for Scientific Research (NWO 0240-001-003); NWO Groot (480-15-001/674): Netherlands Twin Registry Repository: researching the interplay between genome and environment; NWO-Spi-56-464-14192 Biobanking and Biomolecular Resources Research Infrastructure (BBMRI—NL, 184.021.007 and 184.033.111); European Research Council (ERC-230374); the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A); and the National Institute of Mental Health Grand Opportunity grants (grant nos. 1RC2MH089951-01 and 1RC2 MH089995-01). The Texas Twin Project is supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development grants (nos. R01HD083613 and R01HD092548). AddHealth is supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development grant (no. P01HD31921), and GWAS grants (nos. R01HD073342 and R01HD060726), with cooperative funding from 23 other federal agencies and foundations. The Dunedin Multidisciplinary Health and Development Study is supported by the NZ HRC, NZ MBIE, National Institute on Aging grant (no. R01AG032282), and a UK Medical Research Council grant (no. MR/P005918/1). The E-Risk Study is supported by a UK Medical Research Council grant (no. G1002190) and Eunice Kennedy Shriver National Institute of Child Health and Human Development grant (no. R01HD077482). The WLS is supported by National Institute on Aging grants (nos. R01AG041868 and P30AG017266). Some of the work used a high-performance computing facility partially supported by grant no. 2016-IDG-1013 from the North Carolina Biotechnology Center. The Population Research Center at the University of Texas at Austin is supported by NIH grant no. P2CHD042849. P.A.D. is supported by grant no. 531003014 from the Netherlands Organisation for Health Research and Development (ZonMW). P.B. is supported by the NORFACE-DIAL grant no. 462-16-100. S.R.C. is supported by the UK Medical Research Council grant no. MR/R024065/1 and NIH grant no. R01AG054628. E.M.T.-D. is supported by NIH grant nos. R01AG054628 and R01HD083613. A.A. is supported by the Foundation Volksbond Rotterdam and by ZonMw (grant no. 849200011). E.v.B. is supported by NWO VENI (grant no. 451-15-017). D.I.B. acknowledges the Royal Netherlands Academy of Science (KNAW) Professor Award (PAH/6635). B.W.D. is supported by award no. 96-17-04

from the Russell Sage Foundation and the Ford Foundation. H.F.I. was supported by the 'Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies' project (ACTION). ACTION received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no. 602768. J.W. is supported by a postdoctoral fellowship by the AXA Research Fund. D.W.B. is a fellow of the Canadian Institute for Advanced Research Child Brain Development Network. K.P.H. and E.M.T.-D. are Faculty Research Associates of the Population Research Center at the University of Texas at Austin, which is supported by a grant (no. 5-R24-HD042849) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. M.G.N. is supported by ZonMW grants (nos. 849200011 and 531003014), a VENI grant awarded by NWO (VI.Veni.191 G.030) and an NIH grant (no. R01MH120219).

Author contributions

D.W.B., K.P.H., M.G.N., P.A.D. and M.M. conceived and designed the experiment and the idea for the study, with assistance from E.M.T.-D., B.W.D., P.B., C.M. and J.W. P.A.D., M.M., T.T.M., P.B., B.W.D., D.W.B., D.L.C., K.S., S.R.C., M.G.N., A.A. and H.F.I. analyzed the data. D.W.B., K.P.H., M.G.N., M.M., P.A.D. and E.M.T.-D. wrote the paper with helpful contributions from P.B., B.W.D. and S.R.C. A.D.G., L.A., E.v.B., D.I.B., A.C., K.M.H., T.E.M., R.P., J.A.P., B.S.W., E.L.d.Z. and previously mentioned authors contributed to the interpretation of data, provided critical feedback on manuscript drafts and approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-00754-2.

Correspondence and requests for materials should be addressed to D.W.B., K.P.H. or M.G.N.

Peer review information *Nature Genetics* thanks David Cesarini and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Michel Nivard

Last updated by author(s): Sep 17, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Meta-analysis of summary statistics was performed with Metal, release 2011-03-25. The GWAS-by-subtraction and genetic correlation analyses were performed with GenomicSEM v0.0.2, in R 3.4.3. Clumping was performed with Plink 1.9. LD score regressions were done using LDSC v1.0.0. Polygenic scores weights were calculated using LDpred v0.9.09, scores were built with Plink 1.9. Prediction analyses were executed in R for NTR, AddHealth and the Texas Twin Project cohorts, Stata 14 for WLS, and Stata 15 for E-Risk and Dunedin. Bioannotation analyses were done with MAGMA v1.07, stratified LD score regression (LDSC), and FUSION software R package (downloaded on the 22/07/19). Genotyping platform and imputation software for each cohort are described in Supplementary Table 7. Code used to run the analyses is available at: https://github.com/PerlineDemange/non-cognitive A tutorial on how to perform GWAS-by-subtraction: http://rpubs.com/MichelNivard/565885

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

GWAS summary data for NonCog & Cog (excluding 23andMe) are available at: https://www.dropbox.com/s/fry5f82zpfc0r3w/GWAS_sumstats_Cog_NonCog_Demange_et_al_2020806.zip?dl=0
For 23AndMe dataset access, see https://research.23andme.com/dataset-access/.
Part of the National Longitudinal Study of Adolescent to Adult Health (Add Health) data is publicly available and can be downloaded at the following link: https://data.cpc.unc.edu/projects/2/view#public_li. For restricted access data, details of the data sharing agreement and data access requirements can be found at the following link: https://data.cpc.unc.edu/projects/2/view
The Dunedin study datasets reported in the current article are not publicly available due to lack of informed consent and ethical approval, but are available on request by qualified scientists. Requests require a concept paper describing the purpose of data access, ethical approval at the applicant's university, and provision for secure data access. We offer secure access on the Duke, Otago and King's College campuses. All data analysis scripts and results files are available for review. https://moffittcaspi.trinity.duke.edu/research-topics/dunedin
The E-Risk Longitudinal Twin Study datasets reported in the current article are not publicly available due to lack of informed consent and ethical approval, but are available on request by qualified scientists. Requests require a concept paper describing the purpose of data access, ethical approval at the applicant's university, and provision for secure data access. We offer secure access on the Duke and King's College campuses. All data analysis scripts and results files are available for review. https://moffittcaspi.trinity.duke.edu/research-topics/erisk
Netherlands Twin Register data may be accessed, upon approval of the data access committee, email: ntr.datamanagement.fgb@vu.nl.
Researchers will be able to obtain Texas Twins data through managed access. Requests for managed access should be sent to Dr. Elliot Tucker-Drob (tuckerdrob@uexas.edu) and Dr. Paige Harden (harden@utexas.edu), joint principal investigators of the Texas Twin Project.
Wisconsin Longitudinal study data can be requested following this form: https://www.ssc.wisc.edu/wlsresearch/data/Request_Genetic_Data_28_June_2017.pdf

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the Cog and NonCog GWAS, the effective sample sizes were estimated, following Mallard et al. 2020 (Supplementary Note) For each phenotype in the polygenic score analysis, all available data was used, and reported in Supplementary Table 11.
Data exclusions	GWAS are based only on participants with European ancestry. For the PGS analyses, all available data was used, participants with a non-White/European ancestry were excluded, due to the low portability of PGS between different ancestry populations.
Replication	No replication was done. Polygenic score predictions were performed in several cohorts, cohort-specific and meta-analytic estimates are reported. See Supplementary Table 12 for details on which analyses were performed in which cohorts (PGS analyses reported in the manuscript were all performed in 2 to 5 cohorts). Individual cohorts results are plotted in Supplementary Figure 10 and 12 for cognitive performance and personality traits.
Randomization	n/a: We did not use an experimental design
Blinding	n/a: We did not use an experimental design

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Description of the cohorts, including participants population, is available in Supplementary Note, and Supplementary Tables 9-10.
Recruitment	Description of the cohorts, including participants recruitment, is available in Supplementary Note. Several cohorts under-represent individuals with a lower SES background, exclusion of non-European ancestry participants and participants not having provided genetic information is likely to have increased the bias towards higher educated individuals.
Ethics oversight	The National Longitudinal Study of Adolescent to Adult Health received IRB approval from the University of North Carolina. The E-Risk study received ethical approval from Duke University Campus IRB Protocol: 2018-0414B0630 The Environmental Risk Study of Twin Development and Replication Data Studies The Dunedin Study received ethical approval: 17/STH/25/AM05 Health and Disability Ethics Committees Ministry of Health 133 Molesworth Street PO Box 5013 Wellington 0800 4 ETHICS hdecs@moh.govt.nz Ethical approval for NTR was provided by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes 94/105, 96/205, 99/068, 2003/182, 2010/359) and participants provided informed consent. The Texas Twin Project has received IRB approval from the University of Texas at Austin (Protocol Number 2014-11-0021), which was renewed on 11/20/2017. Ethical approval for usage of the Wisconsin Longitudinal Study data was provided by the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at the University of Zurich (OEC IRB # 2018-049).

Note that full information on the approval of the study protocol must also be provided in the manuscript.