



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE  
DELLA RICERCA

## Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

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

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Demange Perline A, Malanchini Margherita, Mallard Travis T, Biroli P, Cox Simon R, Grotzinger Andrew D, et al. (2021). Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. NATURE GENETICS, 7(9), 775-787 [10.1038/s41588-020-00754-2].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/862141> since: 2022-02-21

*Published:*

DOI: <http://doi.org/10.1038/s41588-020-00754-2>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

**Demange, P. A., Malanchini, M., Mallard, T. T., Biroli, P., Cox, S. R., Grotzinger, A. D., ... & Nivard, M. G. (2021). Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics*, 53(1), 35-44.**

The final published version is available online at:

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

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

1 **Investigating the genetic architecture of non-cognitive skills using GWAS-**  
2 **by-subtraction**

3

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

11

12 <sup>1</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

13 <sup>2</sup>Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam,  
14 The Netherlands

15 <sup>3</sup>Research Institute LEARN!, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

16 <sup>4</sup>Department of Biological and Experimental Psychology, Queen Mary University of London,  
17 London, UK.

18 <sup>5</sup>Social, Genetic and Developmental Psychiatric Centre, Institute of Psychiatry, Psychology, &  
19 Neuroscience, King's College London, London, UK.

20 <sup>6</sup>Department of Psychology, University of Texas at Austin, Austin, TX, USA.

21 <sup>7</sup>Department of Economics, University of Zurich, Zurich, Switzerland.

22 <sup>8</sup>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, Edinburgh, UK.

23 <sup>9</sup>Population Research Center, University of Texas at Austin, Austin, TX, USA.

24 <sup>10</sup>Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The  
25 Netherlands.

26 <sup>11</sup>Department of Psychology & Neuroscience, Duke University, Durham, NC, USA.

27 <sup>12</sup>Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham,  
28 NC, USA.

29 <sup>13</sup>Center for Genomic and Computational Biology, Duke University, Durham, NC, USA.

30 <sup>14</sup>Stanford Graduate School of Education, Stanford University, Palo Alto, CA, USA.

31 <sup>15</sup>Department of Sociology and Carolina Population Center, University of North Carolina at Chapel  
32 Hill, Chapel Hill, NC, USA.

33 <sup>16</sup>Institute for Social Research, University of Michigan, Ann Arbor, MI, USA.

34 <sup>17</sup>Department of Psychology and Dunedin Multidisciplinary Health and Development Research Unit,  
35 University of Otago, Dunedin, New Zealand.

36 <sup>18</sup>Department of Epidemiology, Columbia University Mailman School of Public Health, New York,  
37 NY, USA.

38 <sup>19</sup>Robert N. Butler Columbia Aging Center, Columbia University, New York, NY, USA.

39 \*These authors contributed equally to this work.

40 #These authors jointly supervised the work.

41 e-mail: [daniel.belsky@columbia.edu](mailto:daniel.belsky@columbia.edu), [harden@utexas.edu](mailto:harden@utexas.edu), [m.g.nivard@vu.nl](mailto:m.g.nivard@vu.nl)

42

43 **Little is known about the genetic architecture of traits affecting educational attainment**  
44 **other than cognitive ability. We used Genomic Structural Equation Modeling and prior**  
45 **genome-wide association studies (GWAS) of educational attainment ( $n = 1,131,881$ ) and**  
46 **cognitive test performance ( $n = 257,841$ ) to estimate SNP associations with educational**  
47 **attainment variation that is independent of cognitive ability. We identified 157 genome-**  
48 **wide significant loci and a polygenic architecture accounting for 57% of genetic**  
49 **variance in educational attainment. Non-cognitive genetics were enriched in the same**  
50 **brain tissues and cell types as cognitive performance but showed different associations**  
51 **with gray-matter brain volumes. Non-cognitive genetics were further distinguished by**  
52 **associations with personality traits, less risky behavior, and increased risk for certain**  
53 **psychiatric disorders. For socioeconomic success and longevity, non-cognitive and**  
54 **cognitive-performance genetics demonstrated similar-magnitude associations. By**  
55 **conducting a GWAS of a phenotype that was not directly measured, we offer a first**  
56 **view of genetic architecture of non-cognitive skills influencing educational success.**

57

58 “It takes something more than intelligence to act intelligently.”

59 – Fyodor Dostoyevsky, *Crime and Punishment*

60

61 Success in school—and life—depends on skills beyond cognitive ability<sup>1-4</sup>. Randomized  
62 trials of early-life education interventions find substantial benefits to educational outcomes,  
63 employment, and adult health, even though the interventions have no lasting effects on  
64 children’s cognitive functions<sup>5,6</sup>. These results have captured attention of educators and  
65 policy makers, motivating interest in so-called “non-cognitive skills”<sup>7-9</sup>. Non-cognitive skills  
66 suspected to be important for educational success include motivation, curiosity, persistence,

67 and self-control<sup>1,10-13</sup>. However, questions have been raised about the substance of these  
68 skills and the magnitudes of their impacts on life outcomes<sup>14</sup>.

69 Twin studies find evidence that non-cognitive skills are heritable<sup>3,15-18</sup>. Genetic  
70 analysis could help clarify the contribution of these skills to educational attainment and  
71 elucidate their connections with other traits. However, lack of consistent and reliable  
72 measurements of non-cognitive skills in existing genetic datasets pose challenges<sup>19</sup>.

73 To overcome these challenges, we designed a GWAS of a latent trait, *i.e.* a trait not  
74 measured in any of the genotyped subjects<sup>20</sup>. We borrowed the strategy used in the original  
75 analysis of non-cognitive skills within the discipline of economics<sup>21,22</sup>: we defined genetic  
76 influences on non-cognitive skills as the genetic variation in educational attainment that was  
77 not explained by cognitive skills. We then performed GWAS on this residual “non-cognitive”  
78 genetic variation in educational attainment. This approach is a necessarily imperfect  
79 representation of the true relationship between cognitive and non-cognitive skills; in human  
80 development, cognitive abilities and other skills relevant for educational attainment likely  
81 interact dynamically, each influencing the other<sup>23</sup>. Our analysis excludes genetic influences  
82 on education-relevant skills that also influence measured cognitive abilities. The value of this  
83 imperfect approach is to make a quantity otherwise difficult to study tractable for analysis.

84 We conducted analysis using Genomic Structural Equation Modeling (Genomic-  
85 SEM)<sup>24</sup> applied to published GWAS summary statistics for educational attainment and  
86 cognitive performance<sup>25</sup>. Our analysis used these summary statistics to “subtract” genetic  
87 influence on cognitive performance from the association of each single-nucleotide  
88 polymorphism (SNP) with educational attainment. The remaining associations of each SNP  
89 with educational attainment formed a new GWAS of a non-cognitive skills phenotype that  
90 was never directly measured. We call this novel statistical approach GWAS-by-subtraction.

91 We used results from the GWAS-by-subtraction of non-cognitive skills to conduct  
92 two sets of analyses. First, we conducted hypothesis-driven analysis using the phenotypic  
93 annotation approach<sup>26</sup>. We used genetic correlation and polygenic score analysis to test the  
94 hypothesis that non-cognitive skills influence educational and economic attainments and  
95 longevity and to investigate traits and behaviors that constitute non-cognitive skills. Second,  
96 we conducted hypothesis-free bioinformatic annotation analysis to explore the tissues, cell-  
97 types, and brain structures that might distinguish the biology of non-cognitive skills from the  
98 biology mediating cognitive influences on educational attainment.

99

## 100 **Results**

### 101 **GWAS-by-subtraction identifies genetic associations with non-cognitive variance in**

102 **educational attainment.** The term “non-cognitive skills” was originally coined by  
103 economists studying individuals who were equivalent in cognitive ability but who differed in  
104 educational attainment<sup>22</sup>. Our analysis of non-cognitive skills was designed to mirror this  
105 original approach: we focused on genetic variation in educational outcomes not explained by  
106 genetic variation in cognitive ability. Specifically, we applied Genomic Structural Equation  
107 Modeling (Genomic-SEM)<sup>24</sup> to summary statistics from GWASs of educational attainment<sup>25</sup>  
108 and cognitive performance<sup>25</sup>. Both phenotypes were regressed on a latent factor representing  
109 genetic variance in cognitive performance (hereafter “*Cog*”). Educational attainment was  
110 further regressed on a second latent factor representing the residual genetic variance in  
111 educational attainment left over after regressing-out variance related to cognitive  
112 performance (hereafter “*NonCog*”). By construction, *NonCog* genetic variance was  
113 independent of *Cog* genetic variance ( $r_g = 0$ ). In other words, the *NonCog* factor represents  
114 genetic variation in educational attainment that is not accounted for by the *Cog* factor. These  
115 two latent factors were then regressed on individual SNPs, yielding a GWAS of the latent

116 constructs *NonCog* and *Cog*. A graphical representation of the model is presented in **Figure**  
117 **1**. Parameters are derived in terms of the observed moments of the joint distribution of  
118 educational attainment, cognitive performance, and a SNP (see **Supplementary Note**).

119 The *NonCog* latent factor accounted for 57% of total genetic variance in educational  
120 attainment. Using LD Score regression<sup>27</sup>, we estimated SNP-heritability for *NonCog* to be  
121  $h^2_{NonCog} = 0.0637$  ( $SE = 0.0021$ ). After conventional GWAS significance threshold correction,  
122 GWAS of *NonCog* identified 157 independent genome-wide significant lead SNPs  
123 (independent SNPs defined as outside a 250-kb window, or within a 250-kb window and  $r^2 <$   
124 0.1). The results from the *NonCog* GWAS are graphed as a Manhattan plot in **Figure 2**.  
125 *NonCog* and *Cog* GWAS details are reported in **Supplementary Tables 1-4**,  
126 **Supplementary Figure 1**, and the **Supplementary Note**. In addition, we report a series of  
127 sensitivity analyses as follows: analysis of potential biases due to cohort differences  
128 (**Supplementary Table 5** and **Supplementary Figs. 2-4**); analysis of impact of allowing for  
129 positive genetic correlations between *NonCog* and *Cog* (**Supplementary Tables 6** and **7**, and  
130 **Supplementary Figs. 5** and **6**); analysis of impact of allowing for a moderate causal effect of  
131 educational attainment on cognitive performance<sup>28</sup> (**Supplementary Table 8** and  
132 **Supplementary Figs. 7-9**).

133

134 **Phenotypic annotation analysis elucidates behavioral, psychological and psychiatric**  
135 **correlates of non-cognitive skills genetics.** Our phenotypic annotation analyses proceeded  
136 in two steps. First, we conducted polygenic score (PGS) and genetic correlation (rG) analysis  
137 to test whether our GWAS-by-subtraction succeeded in identifying genetic influences that  
138 were important to educational attainment and also distinct from genetic influences on  
139 cognitive ability. Second, we conducted PGS and rG analyses to explore how *NonCog* related



140 to a network of phenotypes that psychology and economics research suggests might form the  
141 basis of non-cognitive influences on educational attainment.

142 *NonCog genetics are distinct from cognitive performance and are important to*  
143 *education, socioeconomic attainment, and longevity.* To establish whether the Genomic-SEM  
144 GWAS-by-subtraction succeeded in isolating genetic variance in education that was  
145 independent of cognitive function, we compared genetic associations of *NonCog* and *Cog*  
146 with educational attainment and cognitive test performance. Results for analysis of education  
147 and cognitive test phenotypes are graphed in **Figure 3**.

148 We conducted PGS analysis of educational attainment in the Netherlands Twin  
149 Register<sup>29</sup> (NTR), National Longitudinal Study of Adolescent to Adult Health<sup>30</sup> (AddHealth),  
150 Dunedin Longitudinal Study<sup>31</sup>, E-Risk<sup>32</sup>, and Wisconsin Longitudinal Study<sup>33</sup> (WLS) cohorts  
151 (meta-analysis  $n = 24,056$ ; cohorts descriptions in **Supplementary Tables 9** and **10** and  
152 **Supplementary Note**). PGS effect-sizes were the same for *NonCog* and *Cog* (*NonCog*  $\beta =$   
153  $0.24$  ( $SE = 0.03$ ), *Cog*  $\beta = 0.24$  ( $SE = 0.02$ ),  $P_{diff} = 0.702$ ; all PGS results are reported in  
154 **Supplementary Tables 11** and **12**). We conducted complementary genetic correlation  
155 analysis using Genomic SEM and GWAS summary statistics from a hold-out-sample GWAS  
156 of educational attainment (**Supplementary Note**). This analysis allowed us to compute an  
157 out-of-sample genetic correlation of *NonCog* with educational attainment. *NonCog* showed a  
158 stronger genetic correlation with educational attainment as compared to *Cog* (*NonCog*  $r_g =$   
159  $0.71$  ( $SE = 0.02$ ), *Cog*  $r_g = 0.57$  ( $SE = 0.02$ ),  $P_{diff} < 0.0001$ ; all genetic correlation results are  
160 reported in **Supplementary Tables 13** and **14**).

161 We conducted PGS analysis of cognitive test performance in the NTR, Texas Twin  
162 Project<sup>34</sup>, Dunedin, E-Risk, and WLS cohorts (combined  $n = 11,351$ ). The goal of our  
163 GWAS-by-subtraction analysis was to exclude, as much as possible, genetic variance in  
164 cognitive ability from genetic variance in skills relevant for education. Consistent with this

165 goal, effect-sizes for *NonCog* PGS associations with full-scale IQ were smaller by half as  
166 compared to *Cog* PGS associations (*NonCog*  $\beta = 0.17$  ( $SE = 0.02$ ), *Cog*  $\beta = 0.29$  ( $SE = 0.03$ );  
167  $P_{diff} < 0.0001$ ). However, the non-zero correlation between the *NonCog* PGS and full-scale  
168 IQ is a reminder that the cognitive performance GWAS used in our GWAS-by-subtraction  
169 analyses does not capture the entirety of genetic influences on all forms of cognitive tests  
170 measured at all points in the lifespan. Additional PGS analyses of IQ subscales are reported  
171 in **Supplementary Figure 10** and **Supplementary Tables 11** and **12**.

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

178 We next examined downstream economic and health outcomes associated with  
179 greater educational attainment<sup>36,37</sup>. In PGS analysis in the AddHealth and Dunedin cohorts ( $n$   
180  $= 6,358$ ), *NonCog* and *Cog* PGSs showed similar associations with occupational attainment  
181 (*NonCog*  $\beta = 0.21$  ( $SE = 0.01$ ), *Cog*  $\beta = 0.21$  ( $SE = 0.01$ ),  $P_{diff} = 0.902$ ). In genetic correlation  
182 analysis, *NonCog* showed a similar relationship to income<sup>38</sup> as *Cog* (*NonCog*  $r_g = 0.62$ , ( $SE =$   
183  $0.04$ ), *Cog*  $r_g = 0.62$  ( $SE = 0.04$ ),  $P_{diff\_fdr} = 0.947$ ) and a stronger relationship with  
184 neighborhood deprivation<sup>38</sup>, a measure related to where a person can afford to live (*NonCog*  
185  $r_g = -0.51$  ( $SE = 0.05$ ), *Cog*  $r_g = -0.32$  ( $SE = 0.04$ ),  $P_{diff\_fdr} = 0.001$ ). In Genomic-SEM  
186 analysis, *NonCog* explained 53% of the genetic correlation between educational attainment  
187 and income and 65% of the genetic correlation between educational attainment and  
188 neighborhood deprivation (**Supplementary Table 15**).

189 We conducted genetic correlation analysis of longevity based on GWAS of parental  
190 lifespan<sup>39</sup>. Genetic correlations were stronger for *NonCog* as compared to *Cog* (*NonCog*  $r_g =$   
191  $0.37$  ( $SE = 0.03$ ); *Cog*  $r_g = 0.27$  ( $SE = 0.03$ );  $P_{diff\_fdr} = 0.024$ ). In Genomic-SEM analysis,  
192 *NonCog* explained 61% of the genetic correlation between educational attainment and  
193 longevity.

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

199 We next conducted a series of genetic correlation analyses to explore the network of  
200 phenotypes to which *NonCog* was genetically correlated. To develop understanding of the  
201 substance of non-cognitive skills, we tested where in that network of phenotypes genetic  
202 correlations with *NonCog* diverged from genetic correlations with *Cog*. Our analysis was  
203 organized around four themes: decision-making preferences, health-risk and fertility  
204 behaviors, personality traits, and psychiatric disorders. Results of genetic correlation analyses  
205 are graphed in **Figure 4** and **Supplementary Figure 11**. Results are reported in  
206 **Supplementary Table 14**.

207 *NonCog* genetics were associated with decision-making preferences. In economics,  
208 non-cognitive influences on achievement and health are often studied in relation to decision-  
209 making preferences<sup>40–43</sup>. *NonCog* was genetically correlated with higher tolerance of risks<sup>44</sup>  
210 ( $r_g = 0.10$  ( $SE = 0.03$ )) and willingness to forego immediate gratification in favor of a larger  
211 reward at a later time<sup>45</sup> (delay discounting  $r_g = -0.52$  ( $SE = 0.08$ )). In contrast, *Cog* was  
212 genetically correlated with generally more cautious decision-making characterized by lower

213 levels of risk tolerance ( $r_g = -0.35$  ( $SE = 0.07$ ),  $P_{\text{diff\_fdr}} < 0.0001$ ) and delay discounting ( $r_g = -$   
214  $0.35$  ( $SE = 0.07$ ),  $P_{\text{diff\_fdr}} = 0.082$ ).

215 *NonCog genetics were associated with less health-risk behavior and delayed fertility.*

216 An alternative approach to studying specific non-cognitive skills is to infer individual  
217 differences in non-cognitive skills from patterns of health-risk behavior. *NonCog* was  
218 genetically correlated with less health-risk behavior as indicated by analysis of obesity<sup>46</sup>,  
219 substance use<sup>44,47-50</sup>, and sexual behaviors and early fertility<sup>44,51,52</sup> ( $r_g$  range 0.2-0.5), with the  
220 exception that the  $r_g$  with alcohol use was not different from zero and  $r_g$  with cannabis use  
221 was positive. Genetic correlations for *Cog* were generally in the same direction but of smaller  
222 magnitude.

223 *NonCog genetics were associated with a broad spectrum of personality*

224 *characteristics linked with social and professional competency.* In psychology, non-cognitive  
225 influences on achievement are conceptualized as personality traits, *i.e.* patterns of stable  
226 individual differences in emotion and behavior. The model of personality that has received  
227 the most attention in genetics is a five-factor model referred to as the Big Five. Genetic  
228 correlation analysis of the Big Five personality traits<sup>53-55</sup> revealed *NonCog* genetics were  
229 most strongly associated with Openness to Experience (being curious and eager to learn;  $r_g =$   
230  $0.30$  ( $SE = 0.04$ )) and were further associated with a pattern of personality characteristic of  
231 changes that occur as people mature in adulthood<sup>56</sup>. Specifically, *NonCog* showed a positive  
232  $r_g$  with Conscientiousness (being industrious and orderly;  $r_g = 0.13$  ( $SE = 0.03$ )), Extraversion  
233 (being enthusiastic and assertive;  $r_g = 0.14$  ( $SE = 0.03$ )), and Agreeableness (being polite and  
234 compassionate;  $r_g = 0.14$  ( $SE = 0.05$ )), and negative  $r_g$  with Neuroticism (being emotionally  
235 volatile;  $r_g = -0.15$  ( $SE = 0.04$ )). Genetic correlations of *Cog* with Openness to Experience  
236 and Neuroticism were similar to those for *NonCog* ( $P_{\text{diff\_fdr-Openness}} = 0.040$ ,  $P_{\text{diff\_fdr-Neuroticism}} =$   
237  $0.470$ ). In contrast, genetic correlations of *Cog* with Conscientiousness, Extraversion, and

238 Agreeableness were in the opposite direction ( $r_g = -0.25$  to  $-0.12$ ,  $P_{\text{diff\_fdr}} < 0.0005$ ). PGS  
239 analysis of personality traits is reported in **Supplementary Table 12, Supplementary**  
240 **Figure 12**, and the **Supplementary Note**.

241 *NonCog genetics were associated with higher risk for multiple psychiatric disorders.*

242 In clinical psychology and psychiatry, research is focused on mental disorders. Mental  
243 disorders are generally associated with impairments in academic achievement and social role  
244 functioning<sup>57,58</sup>. However, positive genetic correlations with educational attainment and  
245 creativity have been reported for some disorders<sup>59,60</sup>. We therefore tested *NonCog*  $r_g$  with  
246 psychiatric disorders based on published case-control GWAS of mental disorders<sup>61–67</sup>.

247 *NonCog* was associated with *higher* risk for multiple clinically defined disorders, including  
248 anorexia nervosa ( $r_g = 0.26$  ( $SE = 0.04$ )), obsessive-compulsive disorder ( $r_g = 0.31$  ( $SE =$   
249  $0.06$ )), bipolar disorder ( $r_g = 0.27$  ( $SE = 0.03$ )), and schizophrenia ( $r_g = 0.26$  ( $SE = 0.02$ )).

250 Genetic correlations between *Cog* and psychiatric disorders were either smaller in magnitude  
251 (anorexia nervosa  $r_g = 0.08$  ( $SE = 0.03$ ),  $P_{\text{diff\_fdr}} < 0.001$ ; obsessive-compulsive disorder  $r_g =$   
252  $0.05$  ( $SE = 0.05$ ),  $P_{\text{diff\_fdr}} = 0.002$ ) or in the opposite direction (bipolar disorder  $r_g = -0.07$  ( $SE$   
253  $= 0.03$ ),  $P_{\text{diff\_fdr}} < 0.001$ ; schizophrenia  $r_g = -0.22$  ( $SE = 0.02$ ),  $P_{\text{diff\_fdr}} < 0.001$ ). Both *NonCog*  
254 and *Cog* showed negative genetic correlations with attention-deficit/hyperactivity disorder  
255 (*NonCog*  $r_g = -0.37$  ( $SE = 0.03$ ), *Cog*  $r_g = -0.37$  ( $SE = 0.04$ ),  $P_{\text{diff\_fdr}} = 0.947$ ).

256 In sum, *NonCog* genetics were associated with phenotypes from economics and  
257 psychology thought to mediate non-cognitive influences on educational success. These  
258 associations contrasted with associations for *Cog* genetics, supporting distinct pathways of  
259 influence on achievement in school and later in life. Opposing patterns of association were  
260 also observed for psychiatric disorders, suggesting that the unexpected positive genetic  
261 correlation between educational attainment and mental health problems uncovered in

262 previous studies<sup>60,68,69</sup> arises from non-cognitive genetic influences on educational  
263 attainment.

264

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

272 *NonCog and Cog genetics were enriched in similar tissues and cells.* We tested  
273 whether common variants in genes specifically expressed in 53 GTEx tissues<sup>70</sup> or in 152  
274 tissues captured in a previous aggregation of RNA-seq studies<sup>71,72</sup> were enriched in their  
275 effects on *Cog* or *NonCog*. Genes predominantly expressed in the brain rather than peripheral  
276 tissues were enriched in both *NonCog* and *Cog* (**Supplementary Table 16**).

277 To examine expression patterns at a more granular level of analysis, we used  
278 MAGMA<sup>73</sup> and stratified LD score regression<sup>74</sup> to test enrichment of common variants in 265  
279 nervous system cell-type-specific gene-sets<sup>75</sup> (**Supplementary Table 17**). In MAGMA  
280 analysis, common variants in 95 of 265 gene-sets were enriched for association with *NonCog*.  
281 The enriched cell-types were predominantly neurons (97%), with enrichment most  
282 pronounced for telencephalon-projecting neurons, di- and mesencephalon neurons, and to a  
283 lesser extent, telencephalon interneurons (**Supplementary Fig. 13** and **Supplementary**  
284 **Table 18**). Enrichment for *Cog* was similar to *NonCog* (correlation between Z-statistics  
285 Pearson's  $r = 0.85$ ), and there were no differences in cell-type-specific enrichment,  
286 suggesting that the same types of brain cells mediate genetic influences on *NonCog* and *Cog*

287 **(Supplementary Fig. 14)**. Stratified LDSC results were similar to results from MAGMA  
288 **(Supplementary Note, Supplementary Fig. 15, and Supplementary Table 19)**.

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

294 *NonCog* and *Cog* genetics show diverging associations with total and regional brain  
295 volumes. Educational attainment has previously been found to be genetically correlated with  
296 greater total brain volume<sup>77,78</sup>. We therefore used a GWAS of regional brain volume to  
297 compare the  $r_g$  of *NonCog* and *Cog* with total brain volume and with 100 regional brain  
298 volumes (99 gray matter volumes and white matter volume) controlling for total brain  
299 volume **(Supplementary Table 21)**<sup>79</sup>. For total brain volume, genetic correlation was  
300 stronger for *Cog* as compared to *NonCog* (*Cog*  $r_g = 0.22$  ( $SE = 0.04$ ), *NonCog*  $r_g = 0.07$  ( $SE =$   
301  $0.03$ ),  $P_{diff} = 0.005$ ). Total gray matter volume, controlling for total brain volume, was not  
302 associated with either *NonCog* or *Cog* (*NonCog*:  $r_g = 0.07$  ( $SE = 0.04$ ); *Cog*:  $r_g = 0.06$  ( $SE =$   
303  $0.04$ )). For total white matter volume, conditional on total brain volume, genetic correlation  
304 was weakly negative for *NonCog* as compared to *Cog* (*NonCog*  $r_g = -0.12$  ( $SE = 0.04$ ), *Cog*  
305 ( $r_g = -0.01$  ( $SE = 0.04$ ),  $P_{diff} = 0.04$ ).

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

311 Finally, we tested genetic correlation of *NonCog* and *Cog* with white matter tract  
312 integrity as measured using diffusion tensor imaging (DTI)<sup>80</sup>. Analyses included 5 DTI  
313 parameters in each of 22 white matter tracts (**Supplementary Table 22**). *NonCog* was  
314 positively associated with the mode of anisotropy parameter (which denotes a more tubular,  
315 as opposed to planar, water diffusion) in the corticospinal tract, retrolenticular limb of the  
316 internal capsule, and splenium of the corpus callosum (**Fig. 5**). However, all correlations  
317 were small ( $0.10 < r_g < 0.14$ ), and we detected no genetic correlations that differed between  
318 *NonCog* and *Cog* (**Supplementary Note**).

319

## 320 **Discussion**

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

330 Phenotypic and biological annotation analyses shed light on the substance of heritable  
331 non-cognitive skills influencing education. Economists hypothesize that preferences that  
332 guide decision-making in the face of risk and delayed rewards represent non-cognitive  
333 influences on educational attainment. Consistent with this hypothesis, *NonCog* genetics were  
334 associated with higher risk tolerance and lower time discounting. These decision-making  
335 preferences are associated with financial wealth, whereas opposite preferences are



336 hypothesized to contribute to a feedback loop perpetuating poverty<sup>81</sup>. Consistent with results  
337 from analysis of decision-making preferences, *NonCog* genetics were also associated with  
338 healthier behavior and later fertility.

339         Psychologists hypothesize that the Big Five personality characteristics of  
340 conscientiousness and openness are the two “pillars of educational success”<sup>2,3,82</sup>. Our results  
341 provide some support for this hypothesis, with the strongest genetic correlation evident for  
342 openness. However, they also show that non-cognitive skills encompass the full range of  
343 personality traits, including agreeableness, extraversion, and the absence of neuroticism. This  
344 pattern mirrors the pattern of personality change that occurs as young people mature into  
345 adulthood<sup>56</sup>. Thus, non-cognitive skills share genetic etiology with what might be termed as  
346 “mature personality”. The absolute magnitudes of genetic correlations between *NonCog* and  
347 individual personality traits are modest. This result suggests that the personality traits  
348 described by psychologists capture some, but not all, genetic influence on non-cognitive  
349 skills.

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

360 manifestations of dimensional psychological traits, which might be associated with adaptive  
361 functioning within the normal range<sup>83-85</sup>.

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

374 We acknowledge limitations. Cognitive and non-cognitive skills develop in  
375 interaction with one another. For example, the dynamic mutualism hypothesis<sup>86</sup> proposes that  
376 non-cognitive characteristics shape investments of time and effort, leading to differences in  
377 the pace of cognitive development<sup>87,88</sup>. However, in Genomic-SEM analysis, the *NonCog*  
378 factor is, by construction, uncorrelated with genetic influences on adult cognition as  
379 measured in the *Cog* GWAS. Our statistical separation of *NonCog* from cognition is thus a  
380 simplified representation of development. Longitudinal studies with repeated measures of  
381 cognitive and candidate non-cognitive skills are needed to study their reciprocal relationships  
382 across development<sup>89,90</sup>. Our statistical separation of *NonCog* from cognition is also  
383 incomplete. The ability to control statistically for any variable, genetic or otherwise, depends  
384 on how well and comprehensively that variable is measured<sup>91</sup>. The tests of cognitive

385 performance included in the *Cog* GWAS likely do not capture all genetic influences on all  
386 forms of cognitive ability across the lifespan<sup>92,93</sup>. Despite these limitations, our simplified and  
387 incomplete statistical separation of *NonCog* from *Cog* allowed us to test whether heritable  
388 traits other than cognitive ability influenced educational attainment and to explore what those  
389 traits might be.

390         Because our analysis was based on GWAS of educational attainment, non-cognitive  
391 genetics identified here may differ from non-cognitive genetics affecting other  
392 socioeconomic attainments like income, or traits and behaviors that mediate responses to  
393 early childhood interventions, to the extent that those genetics do not affect educational  
394 attainment. Parallel analysis of alternative attainment phenotypes will clarify the specificity  
395 of discovered non-cognitive genetics.

396         In the case of GWAS of educational attainment, the included samples were drawn  
397 mainly from Western Europe and the U.S., and participants completed their education in the  
398 late 20<sup>th</sup> and early 21<sup>st</sup> centuries. The phenotype of educational attainment reflects an  
399 interaction between an individual and the social system in which they are educated.  
400 Differences across social systems, including education policy, culture, and historical context,  
401 may result in different heritable traits influencing on educational attainment<sup>94</sup>. Results  
402 therefore may not generalize beyond the times and places GWAS samples were collected.

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

408         Within the bounds of these limitations, results illustrate the application of Genomic-  
409 SEM to conduct GWAS of a phenotype not directly measured in GWAS databases. This

410 application could have broad utility beyond the genetics of educational attainment. The  
411 GWAS-by-subtraction method allowed us to study a previously hard-to-interpret residual  
412 value. Our analysis provides a first view of the genetic architecture of non-cognitive skills  
413 influencing educational success. These skills are central to theories of human capital  
414 formation within the social and behavioral sciences and are increasingly the targets of social  
415 policy interventions. Our results establish that non-cognitive skills are central to the  
416 heritability of educational attainment and illuminate connections between genetic influences  
417 on these skills and social and behavioral science phenotypes.

418

## 419 **Acknowledgements**

420 This study was developed with support from the Jacobs Foundation at a meeting organized  
421 by D.W.B. and K.P.H. with support from E.M.T.-D. and C.M. and also attended by co-  
422 authors P.B., B.W.D., and J.W. We gratefully acknowledge contributions to the meeting from  
423 Katrin Mannik and Felix Tropf, and the Jacobs Foundation Fellowship team who made the  
424 meeting possible. D.W.B., K.P.H., M.G.N., E.M.T.-D., and C.M. are fellows of the  
425 Foundation. J.W. is a Jacobs Foundation Young Scholar.

426 This study used GWAS summary statistics published by the Social Science Genetic  
427 Association Consortium (SSGAC) and additional data obtained from 23andMe. We thank the  
428 research participants and employees of 23andMe for making this work possible. We thank  
429 the SSGAC and COGENT consortia for sharing their summary statistics of the GWASs of  
430 educational attainment and cognitive performance, especially Aysu Okbay for her quick and  
431 repeated help with providing these data. This study used data from the Netherlands Twin  
432 Register (NTR), the Texas Twin Study, the National Longitudinal Study of Adolescent to  
433 Adult Health (Add Health), the Dunedin Longitudinal Study, the E-Risk Study, and the  
434 Wisconsin Longitudinal Study (WLS).

435 NTR is supported by: ‘Twin-family database for behavior genetics and genomics  
436 studies’ (NWO 480-04-004), Longitudinal data collection from teachers of Dutch twins and  
437 their siblings (NWO-481-08-011); Twin-family study of individual differences in school  
438 achievement (NWO 056-32-010) and Gravitation program of the Dutch Ministry of  
439 Education, Culture and Science and the Netherlands Organization for Scientific Research  
440 (NWO 0240-001-003); NWO Groot (480-15-001/674): Netherlands Twin Registry  
441 Repository: researching the interplay between genome and environment; NWO- Spi-56-464-  
442 14192 Biobanking and Biomolecular Resources Research Infrastructure (BBMRI – NL,  
443 184.021.007 and 184.033.111); European Research Council (ERC-230374); the Avera  
444 Institute for Human Genetics, Sioux Falls, South Dakota (USA) and the National Institutes of  
445 Health (NIH, R01D0042157-01A); the NIMH Grand Opportunity grants (1RC2MH089951-  
446 01 and 1RC2 MH089995-01). The Texas Twin Project is supported by Eunice Kennedy  
447 Shriver National Institute of Child Health and Human Development grants R01HD083613  
448 and R01HD092548. Add Health is supported by Eunice Kennedy Shriver National Institute  
449 of Child Health and Human Development grant P01HD31921, and GWAS grants  
450 R01HD073342 and R01HD060726, with cooperative funding from 23 other federal agencies  
451 and foundations. The Dunedin Multidisciplinary Health and Development Study is supported  
452 by the NZ HRC, NZ MBIE, National Institute on Aging grant R01AG032282, and UK  
453 Medical Research Council grant MR/P005918/1. The E-Risk Study is supported by the UK  
454 Medical Research Council grant G1002190 and Eunice Kennedy Shriver National Institute of  
455 Child Health and Human Development grant R01HD077482. The Wisconsin Longitudinal  
456 Study is supported by National Institute on Aging grants R01AG041868 and P30AG017266.  
457 Some of the work used a high-performance computing facility partially supported by  
458 grant 2016-IDG-1013 from the North Carolina Biotechnology Center. The Population

459 Research Center at the University of Texas at Austin is supported by NIH grant  
460 P2CHD042849.

461 P.A.D. is supported by the grant 531003014 from The Netherlands Organisation for  
462 Health Research and Development (ZonMW). P.B. is supported by the NORFACE-DIAL  
463 grant number 462-16-100. S.R.C. is supported by the UK Medical Research Council grant  
464 MR/R024065/1 and NIH grant R01AG054628. EMTD is supported by NIH grants  
465 R01AG054628 and R01HD083613. A.A. is supported by the Foundation Volksbond  
466 Rotterdam and by ZonMw grant 849200011. E.v.B. is supported by NWO VENI grant 451-  
467 15-017. D.I.B. acknowledges the Royal Netherlands Academy of Science (KNAW) Professor  
468 Award (PAH/6635). B.W.D. is supported by award # 96-17-04 from the Russell Sage  
469 Foundation and the Ford Foundation. H.F.I. was supported by the “Aggression in Children:  
470 Unraveling gene-environment interplay to inform Treatment and InterventiON strategies”  
471 project (ACTION). ACTION received funding from the European Union Seventh Framework  
472 Program (FP7/2007-2013) under grant agreement no 602768. J.W. is supported by a  
473 postdoctoral fellowship by the AXA Research Fund. D.W.B. is a fellow of the Canadian  
474 Institute for Advanced Research Child Brain Development Network. K.P.H. and E.M.T.-D.  
475 are Faculty Research Associates of the Population Research Center at the University of Texas  
476 at Austin, which is supported by grant, 5-R24-HD042849, from the Eunice Kennedy Shriver  
477 National Institute of Child Health and Human Development (NICHD). M.G.N. is supported  
478 by ZonMW grants 849200011 and 531003014 from The Netherlands Organisation for Health  
479 Research and Development, a VENI grant awarded by NWO (VI.Veni.191G.030), and NIH  
480 grant R01MH120219.

481

## 482 **Author Contributions**

483 **Conceived and designed the experiment:** D.W.B., K.P.H., M.G.N., P.A.D., and M.M.  
484 conceived the idea for the study with assistance from E.M.T.-D., B.W.D., P.B, C.M., and

485 J.W. **Analyzed the data:** P.A.D., M.M., T.T.M., P.B., B.W.D., D.W.B., D.L.C., K.S.,  
486 S.R.C., M.G.N., A.A., and H.F.I. **Wrote the paper:** D.W.B., K.P.H., M.G.N., M.M., P.A.D.,  
487 and E.M.T.-D. with helpful contributions from P.B., B.W.D., and S.R.C. A.D.G., L.A.,  
488 E.v.B., D.I.B., A.C., K.M.H., T.E.M., R.P., J.A.P., B.S.W., E.L.Z. and previously mentioned  
489 authors contributed to interpretation of data, provided critical feedback on manuscript drafts  
490 and approved the final draft.

491

## 492 **Competing Interests**

493 The authors declare no competing interests.

494

495

496

497 **References**

- 498 1. Moffitt, T. E. *et al.* A gradient of childhood self-control predicts health, wealth, and  
499 public safety. *Proc. Natl. Acad. Sci. USA* **108**, 2693–2698 (2011).
- 500 2. von Stumm, S., Hell, B. & Chamorro-Premuzic, T. The hungry mind: intellectual  
501 curiosity is the third pillar of academic performance. *Perspect. Psychol. Sci.* **6**, 574–588  
502 (2011).
- 503 3. Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D. & Harden, K. P.  
504 Genetically-mediated associations between measures of childhood character and  
505 academic achievement. *J. Pers. Soc. Psychol.* **111**, 790–815 (2016).
- 506 4. Heckman, J. J., Stixrud, J. & Urzua, S. The effects of cognitive and noncognitive abilities  
507 on labor market outcomes and social behavior. *J. Labor Econ.* **24**, 411–482 (2006).
- 508 5. Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. & Yavitz, A. The rate of return to  
509 the HighScope Perry Preschool Program. *J. Public Econ.* **94**, 114–128 (2010).
- 510 6. Conti, G., Heckman, J. J. & Pinto, R. The effects of two influential early childhood  
511 interventions on health and healthy behaviour. *Econ. J.* **126**, F28–F65 (2016).
- 512 7. Gutman, L. M. & Schoon, I. The impact of non-cognitive skills on outcomes for young  
513 people. *Educ. Endow. Found.* **59**, 2019 (2013).
- 514 8. Garcia, E. *The Need to Address Noncognitive Skills in the Education Policy Agenda.*  
515 [https://www.epi.org/publication/the-need-to-address-noncognitive-skills-in-the-](https://www.epi.org/publication/the-need-to-address-noncognitive-skills-in-the-education-policy-agenda/)  
516 [education-policy-agenda/](https://www.epi.org/publication/the-need-to-address-noncognitive-skills-in-the-education-policy-agenda/) (2014).
- 517 9. Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B. & Borghans, L. Fostering and  
518 measuring skills: improving cognitive and non-cognitive skills to promote lifetime  
519 success. OECD Education Working Papers, No. 110, OECD Publishing, Paris. (2014).
- 520 10. Heckman, J. J. Skill formation and the economics of investing in disadvantaged children.  
521 *LIFE CYCLES* **312**, 4 (2006).



- 522 11. Heckman, J. J. & Kautz, T. Hard evidence on soft skills. *Labour Econ.* **19**, 451–464  
523 (2012).
- 524 12. Rimfeld, K., Kovas, Y., Dale, P. S. & Plomin, R. True grit and genetics: Predicting  
525 academic achievement from personality. *J. Pers. Soc. Psychol.* **111**, 780–789 (2016).
- 526 13. Richardson, M., Abraham, C. & Bond, R. Psychological correlates of university students’  
527 academic performance: a systematic review and meta-analysis. *Psychol. Bull.* **138**, 353–  
528 387 (2012).
- 529 14. Smithers, L. G. *et al.* A systematic review and meta-analysis of effects of early life non-  
530 cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat. Hum.*  
531 *Behav.* **2**, 867–880 (2018).
- 532 15. Kovas, Y. *et al.* Why children differ in motivation to learn: Insights from over 13,000  
533 twins from 6 countries. *Personal. Individ. Differ.* **80**, 51–63 (2015).
- 534 16. Loehlin, J. C. *Genes and environment in personality development.* (Sage Publications,  
535 1992).
- 536 17. Tucker-Drob, E. M. & Harden, K. P. Learning motivation mediates gene-by-  
537 socioeconomic status interaction on mathematics achievement in early childhood. *Learn.*  
538 *Individ. Differ.* **22**, 37–45 (2012).
- 539 18. Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P. & Tucker-Drob, E.  
540 M. “Same but different”: associations between multiple aspects of self-regulation,  
541 cognition, and academic abilities. *J. Pers. Soc. Psychol.* **117**, 1164–1188 (2019).
- 542 19. Morris, T. T., Smith, G. D., van Den Berg, G. & Davies, N. M. Investigating the  
543 longitudinal consistency and genetic architecture of non-cognitive skills, and their  
544 relation to educational attainment. <http://biorxiv.org/lookup/doi/10.1101/470682> (2018)  
545 doi:10.1101/470682.

- 546 20. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using  
547 family history of disease. *Nat. Genet.* **49**, 325–331 (2017).
- 548 21. Bowles, S. & Gintis, H. *Schooling In Capitalist America: Educational Reform And The*  
549 *Contradictions Of Economic Life.* (Basic Books, 1977).
- 550 22. Heckman, J. J. & Rubinstein, Y. The importance of noncognitive skills: lessons from the  
551 GED Testing Program. *Am. Econ. Rev.* **91**, 145–149 (2001).
- 552 23. Ackerman, P. L., Kanfer, R. & Goff, M. Cognitive and noncognitive determinants and  
553 consequences of complex skill acquisition. *J. Exp. Psychol. Appl.* **1**, 270–304 (1995).
- 554 24. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the  
555 multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
- 556 25. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association  
557 study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121  
558 (2018).
- 559 26. Belsky, D. W. & Harden, K. P. Phenotypic annotation: using polygenic scores to translate  
560 discoveries from genome-wide association studies from the top down. *Curr. Dir.*  
561 *Psychol. Sci.* **28**, 82–90 (2019).
- 562 27. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from  
563 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 564 28. Ritchie, S. J. & Tucker-Drob, E. M. How much does education improve intelligence? A  
565 meta-analysis. *Psychol. Sci.* **29**, 1358–1369 (2018).
- 566 29. Ligthart, L. *et al.* The Netherlands Twin Register: longitudinal research based on twin  
567 and twin-family designs. *Twin Res. Hum. Genet.* **22**, 623–636 (2019).
- 568 30. Harris, K. M. *et al.* Cohort profile: The National Longitudinal Study of Adolescent to  
569 Adult Health (Add Health). *Int. J. Epidemiol.* **48**, 1415–1415k (2019).

- 570 31. Poulton, R., Moffitt, T. E. & Silva, P. A. The Dunedin Multidisciplinary Health and  
571 Development Study: overview of the first 40 years, with an eye to the future. *Soc.*  
572 *Psychiatry Psychiatr. Epidemiol.* **50**, 679–693 (2015).
- 573 32. Moffitt, T. E. & E-risk Team. Teen-aged mothers in contemporary Britain. *J. Child*  
574 *Psychol. Psychiatry* **43**, 727–742 (2002).
- 575 33. Herd, P., Carr, D. & Roan, C. Cohort profile: Wisconsin longitudinal study (WLS). *Int. J.*  
576 *Epidemiol.* **43**, 34–41 (2014).
- 577 34. Harden, K. P., Tucker-Drob, E. M. & Tackett, J. L. The Texas Twin Project. *Twin Res.*  
578 *Hum. Genet.* **16**, 385–390 (2013).
- 579 35. Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated  
580 with FBNP1L. *Mol. Psychiatry* **19**, 253–258 (2014).
- 581 36. Chetty, R. *et al.* The association between income and life expectancy in the United States,  
582 2001-2014. *JAMA* **315**, 1750–1766 (2016).
- 583 37. Case, A. & Deaton, A. Mortality and morbidity in the 21st century. *Brook. Pap. Econ.*  
584 *Act.* **2017**, 397–476 (2017).
- 585 38. Hill, W. D. *et al.* Molecular genetic contributions to social deprivation and household  
586 income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).
- 587 39. Timmers, P. R. *et al.* Genomics of 1 million parent lifespans implicates novel pathways  
588 and common diseases and distinguishes survival chances. *eLife* **8**, e39856 (2019).
- 589 40. Almlund, M., Duckworth, A. L., Heckman, J. & Kautz, T. Personality psychology and  
590 economics. in *Handbook of the Economics of Education* vol. 4 1–181 (Elsevier, 2011).
- 591 41. Borghans, L., Duckworth, A. L., Heckman, J. J. & Weel, B. ter. The economics and  
592 psychology of personality traits. *J. Hum. Resour.* **43**, 972–1059 (2008).
- 593 42. Rabin, M. A perspective on psychology and economics. *Eur. Econ. Rev.* **29** (2002).

- 594 43. Becker, A., Deckers, T., Dohmen, T., Falk, A. & Kosse, F. The relationship between  
595 economic preferences and psychological personality measures. *Annu. Rev. Econ.* **4**, 453–  
596 478 (2012).
- 597 44. Linnér, R. K. *et al.* Genome-wide association analyses of risk tolerance and risky  
598 behaviors in over 1 million individuals identify hundreds of loci and shared genetic  
599 influences. *Nat. Genet.* **51**, 245–257 (2019).
- 600 45. Sanchez-Roige, S. *et al.* Genome-wide association study of delay discounting in 23,217  
601 adult research participants of European ancestry. *Nat. Neurosci.* **21**, 16–18 (2018).
- 602 46. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body  
603 mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–  
604 3649 (2018).
- 605 47. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci  
606 associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
- 607 48. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common  
608 genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
- 609 49. Schumann, G. *et al.* KLB is associated with alcohol drinking, and its gene product  $\beta$ -  
610 Klotho is necessary for FGF21 regulation of alcohol preference. *Proc. Natl. Acad. Sci.*  
611 *USA* **113**, 14372–14377 (2016).
- 612 50. Pasma, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap  
613 with psychiatric traits, and a causal effect of schizophrenia liability. *Nat. Neurosci.* **21**,  
614 1161–1170 (2018).
- 615 51. Linnér, R. K. *et al.* Multivariate genomic analysis of 1.5 million people identifies genes  
616 related to addiction, antisocial behavior, and health. *bioRxiv* 2020.10.16.342501 (2020)  
617 doi:10.1101/2020.10.16.342501.

- 618 52. Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human  
619 reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
- 620 53. Lo, M.-T. *et al.* Genome-wide analyses for personality traits identify six genomic loci  
621 and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017).
- 622 54. John, O. P., Naumann, L. P. & Soto, C. J. Paradigm shift to the integrative Big Five Trait  
623 taxonomy. *Handb. Personal. Theory Res.* 114–158 (2008) doi:10.1016/S0191-  
624 8869(97)81000-8.
- 625 55. de Moor, M. H. M. *et al.* Meta-analysis of genome-wide association studies for  
626 personality. *Mol. Psychiatry* **17**, 337–349 (2012).
- 627 56. Caspi, A., Roberts, B. W. & Shiner, R. L. Personality development: stability and change.  
628 *Annu. Rev. Psychol.* **56**, 453–484 (2005).
- 629 57. Kessler, R. C. *et al.* Social consequences of psychiatric disorders, I: educational  
630 attainment. *Am. J. Psychiatry* **152**, 1026–1032 (1995).
- 631 58. Breslau, J., Lane, M., Sampson, N. & Kessler, R. C. Mental disorders and subsequent  
632 educational attainment in a US national sample. *J. Psychiatr. Res.* **42**, 708–716 (2008).
- 633 59. Power, R. A. *et al.* Polygenic risk scores for schizophrenia and bipolar disorder predict  
634 creativity. *Nat. Neurosci.* **18**, 953–955 (2015).
- 635 60. Bansal, V. *et al.* Genome-wide association study results for educational attainment aid in  
636 identifying genetic heterogeneity of schizophrenia. *Nat. Commun.* **9**, 3078 (2018).
- 637 61. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine  
638 the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
- 639 62. Ruderfer, D. M. *et al.* Genomic dissection of bipolar disorder and schizophrenia,  
640 including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018).
- 641 63. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies  
642 new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).

- 643 64. Duncan, L. *et al.* Significant locus and metabolic genetic correlations revealed in  
644 genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174**, 850–858  
645 (2017).
- 646 65. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum  
647 disorder. *Nat. Genet.* **51**, 431–444 (2019).
- 648 66. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive–compulsive  
649 disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
- 650 67. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci.  
651 *Nature* **511**, 421–427 (2014).
- 652 68. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.  
653 *Nat. Genet.* **47**, 1236–1241 (2015).
- 654 69. Nieuwboer, H. A., Pool, R., Dolan, C. V., Boomsma, D. I. & Nivard, M. G. GWIS:  
655 genome-wide inferred statistics for functions of multiple phenotypes. *Am. J. Hum. Genet.*  
656 **99**, 917–927 (2016).
- 657 70. The GTEx Consortium *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis:  
658 Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- 659 71. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using  
660 predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
- 661 72. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage  
662 sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
- 663 73. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-  
664 set analysis of GWAS data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
- 665 74. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-  
666 wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

- 667 75. Zeisel, A. *et al.* Molecular architecture of the mouse nervous system. *Cell* **174**, 999-  
668 1014.e22 (2018).
- 669 76. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association  
670 studies. *Nat. Genet.* **48**, 245–252 (2016).
- 671 77. Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W. & Koellinger, P. D. Are bigger  
672 brains smarter? Evidence from a large-scale preregistered study. *Psychol. Sci.* **30**, 43–54  
673 (2019).
- 674 78. Elliott, M. L. *et al.* A polygenic score for higher educational attainment is associated with  
675 larger brains. *Cereb. Cortex* **29**, 3496–3504 (2019).
- 676 79. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants  
677 influencing regional brain volumes and refines their genetic co-architecture with  
678 cognitive and mental health traits. *Nat. Genet.* **51**, 1637–1644 (2019).
- 679 80. Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white matter  
680 microstructure and genetic overlap with cognitive and mental health traits (n = 17,706).  
681 *Mol. Psychiatry*, published online 30 October 2019 (doi: 10.1038/s41380-019-0569-z).
- 682 81. Haushofer, J. & Fehr, E. On the psychology of poverty. *Science* **344**, 862–867 (2014).
- 683 82. Briley, D. A., Domiteaux, M. & Tucker-Drob, E. M. Achievement-relevant personality:  
684 relations with the Big Five and validation of an efficient instrument. *Learn. Individ.*  
685 *Differ.* **32**, 26–39 (2014).
- 686 83. Smoller, J. W. *et al.* Psychiatric genetics and the structure of psychopathology. *Mol.*  
687 *Psychiatry* **24**, 409–420 (2019).
- 688 84. Plomin, R., Haworth, C. M. A. & Davis, O. S. P. Common disorders are quantitative  
689 traits. *Nat. Rev. Genet.* **10**, 872–878 (2009).
- 690 85. Meehl, P. E. Schizotaxia, schizotypy, schizophrenia. *Am. Psychol.* **17**, 827–838 (1962).

- 691 86. von Stumm, S. & Ackerman, P. L. Investment and intellect: a review and meta-analysis.  
692 *Psychol. Bull.* **139**, 841–869 (2013).
- 693 87. Tucker-Drob, E. M. & Harden, K. P. A behavioral genetic perspective on non-cognitive  
694 factors and academic achievement. in *Genetics, Ethics and Education* (eds. Grigorenko,  
695 E. L., Tan, M., Latham, S. R. & Bouregy, S.) 134–158 (Cambridge University Press,  
696 2017). doi:10.1017/9781316340301.007.
- 697 88. Tucker-Drob, E. M. Motivational factors as mechanisms of gene-environment  
698 transactions in cognitive development and academic achievement. in *Handbook of*  
699 *competence and motivation: Theory and application, 2nd ed.* 471–486 (The Guilford  
700 Press, 2017).
- 701 89. Tucker-Drob, E. M. & Harden, K. P. Intellectual interest mediates gene  $\times$  socioeconomic  
702 status interaction on adolescent academic achievement: intellectual interest and G $\times$ E.  
703 *Child Dev.* **83**, 743–757 (2012).
- 704 90. Malanchini, M. *et al.* Reading self-perceived ability, enjoyment and achievement: A  
705 genetically informative study of their reciprocal links over time. *Dev. Psychol.* **53**, 698–  
706 712 (2017).
- 707 91. Westfall, J. & Yarkoni, T. Statistically controlling for confounding constructs is harder  
708 than you think. *PLoS One* **11**, e0152719 (2016).
- 709 92. de la Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M. & Deary, I. J. *Genetic*  
710 *“General Intelligence,” Objectively Determined and Measured.*  
711 <http://biorxiv.org/lookup/doi/10.1101/766600> (2019) doi:10.1101/766600.
- 712 93. Tucker-Drob, E. M. & Briley, D. A. Continuity of genetic and environmental influences  
713 on cognition across the life span: a meta-analysis of longitudinal twin and adoption  
714 studies. *Psychol. Bull.* **140**, 949–979 (2014).



- 715 94. Tropf, F. C. *et al.* Hidden heritability due to heterogeneity across seven populations. *Nat.*  
716 *Hum. Behav.* **1**, 757–765 (2017).
- 717 95. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse  
718 human populations. *Nat. Commun.* **10**, 3328 (2019).
- 719 96. Martin, A. R. *et al.* Human demographic history impacts genetic risk prediction across  
720 diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
- 721
- 722

723 **Figure legends**

724

725 **Figure 1 | GWAS-by-subtraction Genomic-SEM model.** Cholesky model as fitted in  
726 Genomic SEM, with path estimates for a single SNP included as illustration. SNP, cognitive  
727 performance (CP), and educational attainment (EA) are observed variables based on GWAS  
728 summary statistics. The genetic covariance between CP and EA is estimated based on GWAS  
729 summary statistics for CP and EA. The model is fitted to a 3 x 3 observed variance-  
730 covariance matrix (i.e. SNP, CP, EA). *Cog* and *NonCog* are latent (unobserved) variables.  
731 The covariances between CP and EA and between *Cog* and *NonCog* are fixed to 0. The  
732 variance of the SNP is fixed to the value of  $2pq$  ( $p$  = reference allele frequency,  $q$  =  
733 alternative allele frequency, based on 1000 Genomes phase 3). The residual variances of CP  
734 and EA are fixed to 0, so that all variance is explained by the latent factors. The variances of  
735 the latent factors are fixed to 1. The observed variables CP and EA were regressed on the  
736 latent variables resulting in the estimates for the path loadings:  $\lambda_{\text{Cog-CP}} = 0.4465$ ;  $\lambda_{\text{Cog-EA}}$   
737  $= 0.2237$ ;  $\lambda_{\text{NonCog-EA}} = 0.2565$ . The latent variables were then regressed on each SNP that  
738 met QC criteria.

739

740 **Figure 2 | Manhattan plot of SNP associations with *NonCog*.** Plot of the  $-\log_{10}(P\text{-value})$   
741 associated with the Wald test (two-sided) of  $\beta_{\text{NonCog}}$  for all SNPs, ordered by chromosome  
742 and base position. Purple triangles indicate genome-wide significant ( $P < 5 \times 10^{-8}$ ) and  
743 independent (within a 250-kb window and  $r^2 < 0.1$ ) associations. The red dashed line marks

744 the threshold for genome-wide significance ( $P = 5 \times 10^{-8}$ ), and the black dashed line the  
745 threshold for nominal significance ( $P = 1 \times 10^{-5}$ ).

746

747 **Figure 3 | Polygenic prediction and genetic correlations with IQ and educational**

748 **achievement. a**, Genetic correlations of *NonCog* and *Cog* with educational attainment,

749 highest math class taken, self-reported math ability, and childhood IQ. The dots represent

750 genetic correlations estimated using Genomic SEM. Correlations with *NonCog* are in orange,

751 and with *Cog* in blue. Error bars represent 95% CIs. Exact estimates and *P*-values are

752 reported in **Supplementary Table 14**. For analysis of genetic correlations with educational

753 attainment, we re-ran the Genomic-SEM model to compute *NonCog* and *Cog* using summary

754 statistics that omitted the 23andMe sample from the educational attainment GWAS. We then

755 used the 23andMe sample to run the GWAS of educational attainment. Thus, there is no

756 sample overlap in this analysis. **b**, Effect-size distributions from meta-analysis of *NonCog*

757 and *Cog* polygenic score associations with cognitive test performance and educational

758 attainment. Outcomes were regressed simultaneously on *NonCog* and *Cog* polygenic scores.

759 Effect-sizes entered into the meta-analysis were standardized regression coefficients

760 interpretable as Pearson *r*. Exact estimates and *P*-values are reported in **Supplementary**

761 **Table 12**. Samples and measures are detailed in **Supplementary Tables 9 and 10**. Traits

762 were measured in different samples: educational attainment was measured in the AddHealth,

763 Dunedin, E-Risk, NTR and WLS samples ( $n = 24,056$ ); reading achievement and

764 mathematics achievement were measured in the AddHealth, NTR, and Texas-Twin samples

765 ( $n = 9,274$  for reading achievement;  $n = 10,747$  for mathematics achievement); cognitive test

766 performance (IQ) was measured in the Dunedin, E-Risk, NTR, Texas Twins and WLS

767 samples ( $n = 11,351$ ). The densities were obtained by randomly generating normal

768 distributions where the meta-analytic estimate was included as the mean and the meta-  
769 analytic standard error as the standard deviation.

770

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

782

783 **Figure 5 | Genetic correlations with regional gray matter volumes and white matter**  
784 **tracts. a,** Cortical patterning of FDR-corrected significant genetic correlations with regional  
785 gray matter volumes for *Cog* versus *NonCog*, after correction for total brain volume. Regions  
786 of interest are plotted according to the Desikan-Killiany-Tourville atlas<sup>102</sup>, shown on a single  
787 manually-edited surface (<http://mindboggle.info><sup>103</sup>). Exact estimates and  $P$ -values are  
788 reported in **Supplementary Table 21**. *Cog* showed significant associations with gray matter  
789 volume for the bilateral fusiform, insula and posterior cingulate, the left superior temporal  
790 and left pericalcarine and right superior parietal volumes. *NonCog* was not associated with  
791 any of the regional brain volumes. **b,** White matter tract patterning of FDR-corrected  
792 significant genetic correlations with regional mode of anisotropy (MO) for *Cog* versus

793 *NonCog*. White matter tract probability maps are plotted according to the Johns Hopkins  
794 University DTI atlas (<https://identifiers.org/neurovault.image:1401>)<sup>104</sup>. Exact estimates and  
795 *P*-values are reported in **Supplementary Table 21**. *Cog* was not associated with regional  
796 *MO*. *NonCog* showed significant associations with *MO* in the corticospinal tract, the  
797 retrolenticular limb of the internal capsule and the splenium of the corpus callosum.

798

799

800

801

## 802 **Methods**

803 **Meta-analysis of educational attainment GWAS.** We reproduced the Social Science  
804 Genetic Association Consortium (SSGAC) 2018 GWAS of educational attainment<sup>25</sup> by meta-  
805 analyzing published summary statistics for  $n = 766,345$  (www.thessgac.org/data) with  
806 summary statistics obtained from 23andMe, Inc. ( $n = 365,538$ ). We included SNPs with  
807 sample size  $> 500,000$  and MAF  $> 0.005$  in the 1000 Genomes reference set (10,101,243  
808 SNPs). We did not apply genomic control, as standard errors of publicly available and  
809 23andMe summary statistics were already corrected<sup>25</sup>. Meta-analysis was performed using  
810 METAL<sup>97</sup>.

811

812 **GWAS-by-subtraction.** The objective of our GWAS-by-subtraction analysis was to  
813 estimate, for each SNP, the association with educational attainment that was independent of  
814 that SNP's association with cognition (hereafter, the *NonCog* SNP effect). We used  
815 Genomic-SEM<sup>24</sup> in R 3.4.3 to analyze GWAS summary statistics for the educational  
816 attainment and cognitive performance phenotypes in the SSGAC's 2018 GWAS<sup>25</sup>. The  
817 model regressed the educational-attainment and cognitive-performance summary statistics on  
818 two latent variables, *Cog* and *NonCog* (**Fig. 1**). *Cog* and *NonCog* were then regressed on each  
819 SNP in the genome. This analysis allowed for two paths of association with educational  
820 attainment for each SNP. One path was fully mediated by *Cog*. The other path was  
821 independent of *Cog* and measured the non-cognitive SNP effect, *NonCog*. To identify  
822 independent hits with  $P < 5 \times 10^{-8}$  (the customary  $P$ -value threshold to approximate an alpha  
823 value of 0.05 in GWAS), we pruned the results using a radius of 250 kb and an LD threshold  
824 of  $r^2 < 0.1$  (**Supplementary Tables 1-3**). We explore alternative lead SNPs and loci  
825 definition in **Supplementary Table 4**. The parameters estimated in a GWAS-by-subtraction  
826 and their derivation in terms of the genetic covariance are described in the **Supplementary**

827 **Note** (model specification), and practical analysis steps are further described in the  
828 **Supplementary Note** (SNP filtering). The effective sample size of the *NonCog* and *Cog*  
829 GWAS was estimated to 510,795 and 257,700, respectively (see **Supplementary Note**). We  
830 investigated biases from unaccounted-for heterogeneity in overlap across SNPs in the  
831 educational attainment and cognitive performance GWAS and describe possible strategy to  
832 deal with it (**Supplementary Note**). We investigated potential biases due to cohort  
833 differences in SNP heritability in the **Supplementary Note**. We evaluated the consequences  
834 of modifying  $r_g(\text{NonCog}, \text{Cog}) = 0$  by evaluating  $r_g = 0.1, 0.2$  or  $0.3$ , and we investigated the  
835 consequences of a violation of the assumed causation between cognitive performance and  
836 educational attainment in the **Supplementary Note**.

837

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

849

850 **Polygenic score analysis.** Polygenic score analyses were conducted in data drawn from six  
851 population-based cohorts from the Netherlands, the U.K., the U.S., and New Zealand: (1) the

852 Netherlands Twin Register (NTR)<sup>29,98</sup>, (2) E-Risk<sup>32</sup>, (3) the Texas Twin Project<sup>34</sup>, (4) the  
853 National Longitudinal Study of Adolescent to Adult Health (AddHealth)<sup>30,99</sup>, dbGaP  
854 accession phs001367.v1.p1; (5) Wisconsin Longitudinal Study on Aging (WLS)<sup>33</sup>, dbGaP  
855 accession phs001157.v1.p1; and (6) the Dunedin Multidisciplinary Health and Development  
856 Study<sup>31</sup>. **Supplementary Tables 9 and 10** describe cohort-specific metrics, and we include a  
857 short description of the cohorts' populations and recruitment in **Supplementary Note**. Only  
858 participants with European ancestry were included in the analysis, due to the low portability  
859 of PGS between different ancestry populations. Polygenic scores were computed with PLINK  
860 based on weights derived using the LD-pred<sup>100</sup> software with an infinitesimal prior and the  
861 1000 Genomes phase 3 sample as a reference for the LD structure. LD-pred weights were  
862 computed in a shared pipeline to ensure comparability between cohorts. Each outcome (*e.g.*,  
863 IQ score) was regressed on the *Cog* and *NonCog* polygenic scores and a set of control  
864 variables (sex, 10 principal components derived from the genetic data and, for cohorts in  
865 which these quantities varied, genotyping chip and age), using Stata 14 for WLS, Stata 15 for  
866 E-Risk and the Dunedin Study, and R (versions 3.4.3 and newer) for NTR, AddHealth, and  
867 the Texas Twin Project. In cohorts containing related individuals, non-independence of  
868 observations from relatives was accounted for using generalized estimation equations (GEE)  
869 or by clustering of standard errors at the family level. We used a random effects meta-  
870 analysis to aggregate the results across the cohorts. This analysis allows a cohort-specific  
871 random intercept. Individual cohort results are in **Supplementary Table 11** and meta-  
872 analytic estimates in **Supplementary Table 12**.

873

874 **Biological annotation.** *Enrichment of tissue-specific gene expression.* We used gene-sets  
875 defined in Finucane et al.<sup>101</sup> to test for the enrichment of genes specifically expressed in one  
876 of 53 GTEx tissues<sup>70</sup>, or 152 tissues captured by the Franke et al. aggregation of RNA-seq



877 studies<sup>71,72</sup>. This analysis seeks to confirm the role of brain tissues in mediating *Cog* and  
878 *NonCog* influences on educational attainment. The exact analysis pipeline used is available  
879 online (<https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses>).

880 *Enrichment of cell-type specific expression.* We leveraged single cell RNA  
881 sequencing (scRNA-seq) data of cells sampled from the mouse nervous system<sup>75</sup> to identify  
882 cell-type specific RNA expression. Zeisel et al.<sup>75</sup> sequenced cells obtained from 19 regions in  
883 the contiguous anatomical regions in the peripheral sensory, enteric, and sympathetic nervous  
884 system. After initial QC, they retained 492,949 cells, which were sampled down to 160,796  
885 high quality cells. These cells were further grouped into clusters representing 265 broad cell-  
886 types. We analyzed the dataset published by Zeisel et al. containing mean transcript counts  
887 for all genes with count >1 for each of the 265 clusters (**Supplementary Table 17**). We  
888 restricted analysis to genes with expression levels above the 25<sup>th</sup> percentile. For each gene in  
889 each cell-type, we computed the cell-type specific proportion of reads for the gene  
890 (normalizing the expression within cell-type). We then computed the proportion of  
891 proportions over the 265 cell-types (computing the specificity of the gene to a specific cell-  
892 type). We ranked the 12,119 genes retained in terms of specificity to each cell-type and then  
893 retained the 10% of genes most specific to a cell-type as the “cell-type specific” gene-set. We  
894 then tested whether any of the 265 cell-type specific gene-sets were enriched in the *Cog* or  
895 *NonCog* GWAS. This analysis sought to identify specific cell-types and specific regions in  
896 the brain involved in the etiology of *Cog* and *NonCog*. We further computed the difference in  
897 enrichment for *Cog* and *NonCog* to test whether any cell types were specific to either trait.  
898 For these analyses, we leveraged two widely used enrichment analysis tools: MAGMA<sup>73</sup> and  
899 stratified LD score regression<sup>74</sup> with the European reference panel from 1000 Genomes  
900 Project Phase 3 as SNP location and LD structure reference, Gencode release 19 as gene

901 location reference and the human-mouse homology reference from MGI  
902 ([http://www.informatics.jax.org/downloads/reports/HOM\\_MouseHumanSequence.rpt](http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt)).

903 *MAGMA*. We used *MAGMA* (v1.07b<sup>73</sup>), a program for gene-set analysis based on  
904 GWAS summary statistics. We computed gene-level association statistics using a window of  
905 10 kb around the gene for both *Cog* and *NonCog*. We then used *MAGMA* to run a  
906 competitive gene-set analysis, using the gene *P*-values and gene correlation matrix (reflecting  
907 LD structure) produced in the gene-level analysis. The competitive gene-set analysis tests  
908 whether the genes within the cell-type-specific gene-set described above are more strongly  
909 associated with *Cog/NonCog* than other genes.

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

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

920

$$921 \quad Z_{\text{diff}} = \frac{e_{\text{Cog}} - e_{\text{NonCog}}}{\sqrt{se_{\text{Cog}}^2 + se_{\text{NonCog}}^2 - 2 * CTI * se_{\text{Cog}} * se_{\text{NonCog}}}} \quad (\text{Equation 1})$$

922

923 where  $e_{\text{Cog}}$  is the enrichment of a particular gene-set for *Cog*,  $e_{\text{NonCog}}$  is the enrichment for  
924 the same gene-set for *NonCog*,  $se_{\text{Cog}}$  is the standard error of the enrichment for *Cog*,

925  $se_{NonCog}$  is the standard error of the enrichment for *NonCog*, and CTI is the LD score cross-  
926 trait intercept, a metric of dependence between the GWASs of *Cog* and *NonCog*.

927 We investigated the significance of the difference between *Cog* and *NonCog* tau  
928 coefficient with Equation 1 as well as by computing jackknifed standard errors. From the  
929 jackknifed estimates of the coefficient output by the LDSC software, we computed the  
930 jackknifed estimates and standard errors of the difference between *Cog* and *NonCog* tau  
931 coefficients, as well as a  $z$ -statistic for each annotation.

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

940

941 **Brain modalities. Brain volumes.** We conducted genetic correlation analysis of brain  
942 volumes using GWAS results published by Zhao et al.<sup>79</sup>, who performed GWAS of total  
943 brain volume and 100 regional brain volumes, including 99 gray matter volumes and total  
944 white matter volume (**Supplementary Table 21**). Analyses included covariate adjustment for  
945 sex, age, their square interaction and 20 principle components. Analyses of regional brain  
946 volumes additionally included covariate adjustment for total brain volume. GWAS summary  
947 statistics for these 101 brain volumes were obtained from  
948 <https://med.sites.unc.edu/big2/data/gwas-summary-statistics/>. Summary statistics were  
949 filtered and pre-processed using Genomic-SEM's "munge" function, retaining all HapMap3

950 SNPs with allele frequency  $> 0.01$  outside the MHC region. We used Genomic-SEM to  
951 compute the genetic correlations between *Cog*, *NonCog* and brain volumes. Analyses of  
952 regional volumes controlled for total brain volume. For each volume, we tested whether  
953 correlations differed between *Cog* and *NonCog*. Specifically, we used a chi-squared test to  
954 evaluate the null hypothesis that the two genetic correlations were equal. We used FDR  
955 adjustment to correct for multiple testing. The FDR adjustment is applied to the results for all  
956 gray matter volumes for *Cog* and *NonCog* separately.

957 *White matter structures.* We conducted genetic-correlation analysis of white-matter  
958 structures using GWAS results published by Zhao et al.<sup>80</sup>, who performed GWAS of  
959 diffusion tensor imaging (DTI) measures of the integrity of white-matter tracts. DTI  
960 parameters were derived for fractional anisotropy (FA), mean diffusivity (MD), axial  
961 diffusivity (AD), radial diffusivity (RD), and mode of anisotropy (MO). Each of these  
962 parameters was measured for 22 white matter tracts of interests (**Supplementary Table 22**),  
963 resulting in 110 GWAS. GWAS summary statistics for these 110 GWAS were obtained from  
964 <https://med.sites.unc.edu/bigs2/data/gwas-summary-statistics/>. Summary statistics were  
965 filtered and processed using Genomic-SEM's "munge" function, retaining all HapMap3  
966 SNPs with allele frequency  $> 0.01$  outside the MHC region. For each white matter structure,  
967 we tested whether genetic correlations differed between *Cog* and *NonCog*. Specifically, we  
968 used a chi-squared test to evaluate the null hypothesis that the two genetic correlations were  
969 equal. We used FDR adjustment to correct for multiple testing. As these different diffusion  
970 parameters are statistically and logically interdependent, having been derived from the same  
971 tensor, FDR adjustment was applied to the results for each type of white matter diffusion  
972 parameter separately. FDR correction was applied separately for *Cog* and *NonCog*.

973

974 **Additional Resources**

975 A FAQ on why, how and what we studied is available here:

976 [https://medium.com/@kph3k/investigating-the-genetic-architecture-of-non-cognitive-skills-](https://medium.com/@kph3k/investigating-the-genetic-architecture-of-non-cognitive-skills-using-gwas-by-subtraction-b8743773ce44)  
977 [using-gwas-by-subtraction-b8743773ce44](https://medium.com/@kph3k/investigating-the-genetic-architecture-of-non-cognitive-skills-using-gwas-by-subtraction-b8743773ce44)

978 A tutorial on how to perform GWAS-by-subtraction: <http://rpubs.com/MichelNivard/565885>

979 Additional resources to Genomic SEM software:

980 - A wiki including numerous tutorials:

981 <https://github.com/MichelNivard/GenomicSEM/wiki>

982 - A Genomic SEM user group for specific questions relating to models and  
983 software: <https://groups.google.com/g/genomic-sem-users>

984 - A venue to report technical issues:

985 <https://github.com/MichelNivard/GenomicSEM/issues>

986

## 987 **Code availability**

988 Code used to run the analyses is available at: [https://github.com/PerlineDemange/non-](https://github.com/PerlineDemange/non-cognitive)  
989 [cognitive](https://github.com/PerlineDemange/non-cognitive)

990 A tutorial on how to perform GWAS-by-subtraction: <http://rpubs.com/MichelNivard/565885>

991 All additional software used to perform these analyses are available online.

992

## 993 **Data availability**

994 GWAS summary data for *NonCog* and *Cog* (excluding 23andMe) have been deposited in the

995 GWAS Catalog with accession numbers GCST90011874 and GCST90011875, respectively

996 (*NonCog* GWAS: [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/GCST90011874](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011874),

997 *Cog* GWAS: [ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/GCST90011875](ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011875)).

998

999 For 23andMe dataset access, see <https://research.23andme.com/dataset-access/>.

1000 Part of the National Longitudinal Study of Adolescent to Adult Health (Add Health) data is  
1001 publicly available and can be downloaded at the following link:  
1002 [https://data.cpc.unc.edu/projects/2/view#public\\_li](https://data.cpc.unc.edu/projects/2/view#public_li). For restricted access data, details of the  
1003 data sharing agreement and data access requirements can be found at the following link:  
1004 <https://data.cpc.unc.edu/projects/2/view>

1005 The Dunedin study datasets reported in the current article are not publicly available due to  
1006 lack of informed consent and ethical approval, but are available on request by qualified  
1007 scientists. Requests require a concept paper describing the purpose of data access, ethical  
1008 approval at the applicant's university, and provision for secure data access. We offer secure  
1009 access on the Duke, Otago and King's College campuses. All data analysis scripts and results  
1010 files are available for review (<https://moffittcaspi.trinity.duke.edu/research-topics/dunedin>).

1011 The E-Risk Longitudinal Twin Study datasets reported in the current article are not publicly  
1012 available due to lack of informed consent and ethical approval, but are available on request  
1013 by qualified scientists. Requests require a concept paper describing the purpose of data  
1014 access, ethical approval at the applicant's university, and provision for secure data access.  
1015 We offer secure access on the Duke and King's College campuses. All data analysis scripts  
1016 and results files are available for review ([https://moffittcaspi.trinity.duke.edu/research-](https://moffittcaspi.trinity.duke.edu/research-topics/erisk)  
1017 [topics/erisk](https://moffittcaspi.trinity.duke.edu/research-topics/erisk)).

1018 Netherlands Twin Register data may be accessed, upon approval of the data access  
1019 committee (email: [ntr.datamanagement.fgb@vu.nl](mailto:ntr.datamanagement.fgb@vu.nl)).

1020 Researchers will be able to obtain Texas Twins data through managed access. Requests for  
1021 managed access should be sent to Dr. Elliot Tucker-Drob ([tuckerdrob@utexas.edu](mailto:tuckerdrob@utexas.edu)) and Dr.  
1022 Paige Harden ([harden@utexas.edu](mailto:harden@utexas.edu)), joint principal investigators of the Texas Twin Project.

1023 Wisconsin Longitudinal study data can be requested following this form:  
1024 [https://www.ssc.wisc.edu/wlsresearch/data/Request\\_Genetic\\_Data\\_28\\_June\\_2017.pdf](https://www.ssc.wisc.edu/wlsresearch/data/Request_Genetic_Data_28_June_2017.pdf)

1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050

## METHODS-ONLY REFERENCES

97. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
98. 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).
99. Highland, H. M., Avery, C. L., Duan, Q., Li, Y. & Harris, K. M. *Quality control analysis of Add Health GWAS data.*  
[https://www.cpc.unc.edu/projects/addhealth/documentation/guides/AH\\_GWAS\\_QC.pdf](https://www.cpc.unc.edu/projects/addhealth/documentation/guides/AH_GWAS_QC.pdf)  
(2018).
100. Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
101. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
102. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 171 (2012).
103. Klein, A. Mindboggle-101 manually labeled individual brains. (2016)  
doi:10.7910/DVN/HMQKCK.
104. Gorgolewski, K. J. *et al.* NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinformatics* **9**, 8 (2015).