Polygenic Scores for Cognitive Abilities and their Association with Different Aspects of General Intelligence – a Deep Phenotyping Approach

Erhan Genç * [°] [a,b], Caroline Schlüter [°] [a], Christoph Fraenz [a], Larissa Arning [c], Huu Phuc Nguyen [c], Manuel C. Voelkle [d], Fabian Streit [e], Onur Güntürkün [a], Robert Kumsta [f], Sebastian Ocklenburg [a]

[a]: Department of Biopsychology, Institute of Cognitive Neuroscience, Ruhr University Bochum

[b]: Department of Psychology and Neurosciences, Leibniz Research Centre for Working Environment and Human Factors (IfADo), Dortmund

[c]: Department of Human Genetics, Ruhr University Bochum

[d]: Department of Psychological Research Methods, Institute of Psychology, Humboldt University Berlin

[e]: Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg

[f]: Department of Genetic Psychology, Ruhr University Bochum

[°]: These authors contributed equally

*: Corresponding Author and Lead Contact

Dr. Erhan Genç

Leibniz Research Centre for Working Environment and Human Factors (IfADo)

Department of Psychology and Neurosciences, MRI Unit

Ardeystraße 67

44139 Dortmund

Tel: +49 231 1084203

E-Mail: genc@ifado.de

Abstract

Intelligence is a highly polygenic trait and GWAS have identified thousands of DNA variants contributing with small effects. Polygenic scores (PGS) can aggregate those effects for trait prediction in independent samples. As large-scale light-phenotyping GWAS operationalized intelligence as performance in rather superficial tests, the question arises which intelligence facets are actually captured. We used deep-phenotyping to investigate the molecular determinantes of individual differences in cognitive ability. We therefore studied the association between PGS of educational attainment (EA-PGS) and intelligence (IQ-PGS) with a wide range of intelligence facets in a sample of 320 healthy adults. EA-PGS and IQ-PGS had the highest incremental R^2s for general (3.25%; 1.78%), verbal (2.55%; 2.39%) and numerical intelligence (2.79%; 1.54%) and the weakest for non-verbal intelligence (0.50%; 0.19%) and short-term memory (0.34%; 0.22%). These results indicate that PGS derived from light-phenotyping GWAS do not reflect different facets of intelligence equally well, and thus should not be interpreted as genetic indicators of intelligence per se. The findings refine our understanding of how PGS are related to other traits or life outcomes.

Keywords: Polygenic scores, intelligence, educational attainment, deep-phenotyping, cognitive abilities

Introduction

Gaining insight into the molecular determinants of differences in cognitive abilities is one of the core aims of neurobiological intelligence research. Our ability to "understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning [and] to overcome obstacles by taking thought" has usually been described as general intelligence [1]. Various tests have been designed to measure the cognitive abilities of a person by assessing different aspects of general intelligence, like inductive and deductive reasoning abilities or the amount of acquired declarative knowledge [2]. Whereas most test procedures cover different aspects of general intelligence, there are also tests that focus on specific cognitive abilities. Matrix reasoning tests, for insctance, are typically used to assess non-verbal abstract reasoning [3,4], while other tests measure stored long-term memory of static information like rules, relationships, abstract concepts, and of course, general knowledge [2,5].

Decades of intelligence research have shown that general intelligence is one of the best predictors of important life outcomes, including educational and occupational success [6,7] as well as mental and physical health [8-10]. Thus, considerable research efforts have been put into place to explore the mechanisms behind interindividual differences in general intelligence. Behavioral genetics has been particularly fruitful for intelligence research. Twin and family studies have demonstrated that general intelligence is one of the most heritable behavioral traits, with heritability estimates ranging from 60% to 80% in adulthood [for an overview see 11,12]. However, these traditional quantitative genetics studies cannot be used to estimate which and how many genetic variants contribute to heritability. Although, as early as 1918, Fisher's infinitesimal model postulated that complex traits are affected by a large number of genes, it was not until the advent of genome-wide association studies (GWAS) that effect sizes of single nucleotide polymorphisms (SNPs) could be systematically assessed over the genome. In support of Fisher's infinitesimal model [13], large GWAS on intelligence have shown that even the most strongly associated SNPs explain less than 1% of the variance, and that heritability of intelligence is caused by a very large number of DNA variants of small effect (not taking into account rare mutations with potentially large effects on individuals, but small effects on the population) [14].

Given that GWAS of sufficient sample size can reliably detect very small effects of single common variants, and given that SNPs contribute cumulatively to heritability, a fruitful approach forward has been the use of so-called polygenic scores (PGS). These are genetic indices of a trait, defined as the sum of trait-associated alleles across many genetic loci, weighted by effect sizes estimated by GWAS. Such scores can be calculated for individuals in target samples (independent from the initial discovery GWAS) and be used to predict traits of

interest [15]. For instance, PGS for intelligence (IQ-PGS) [16,17] and educational attainment (EA-PGS) [18-20], a secondary measure of intelligence, have been associated with a wide variety of traits, including life-course development, educational achievement, body mass index or emotional and behavioral problems in children [21-23]. Although IQ-PGS and EA-PGS explain a considerable amount of variance in intelligence (which is thought to increase even further with larger GWAS) [14], and robust and sometimes unexpected associations between genetic indices of cognitive abilities and other traits have been uncovered, it is important to understand that the predictive power of these PGS depends on the cognitive measure that is being used. To reliably identify genetic variants associated with a complex continuous behavioral trait, such as intelligence, in a GWAS, large samples sizes in the 100,000s to millions are required. This has been successfully achieved using a light-phenotyping approach, that is, performing GWAS on the performance in rather superficial tests of general cognitive abilities [24], or even more crudely, years of education [25]. The question thus arises, which of the various aspects of general intelligence [2,1], are mainly reflected in those GWAS. The study at hand aimed to tackle this issue by pursuing a deep phenotyping approach. Using an extensive test battery comprised of tests for memory, processing speed, reasoning, and general knowledge, we investigated the predictive power of EA-PGS [25] and IQ-PGS [24] with regard to each of the aforementioned cognitive abilities.

Methods

Sample size estimation

A statistical power analysis was performed for sample size estimation based on data reported by Savage et al. [24] and Lee et al. [25]. Both meta-analyses reported effect sizes [26] of R^2 = .052 (IQ-PSG) [24], and R^2 =.097 (EA-PGS) [25], respectively. G-power [27] was used to determine the sample size required to detect a small to medium effect size (f^2 = .08) in a multiple linear regression analysis, using an α error of .05 and a statistical power of 1-β = .90. Sample size estimation predicted that *N* = 236 participants were needed to obtain the desired statistical power. With a final sample size of *N*=296 (see below), our sample was thus adequately powered for the main objective of this study.

Participants

We investigated 320 neurologically and psychologically healthy participants with a mean age of 27.68 years (*SD* = 10.51; range 18 - 75 years), including 167 males (mean age: 28.38 years, *SD* = 10.85 years) and 153 females (mean age: 26.92 years, *SD* = 10.11 years). The sample was mainly comprised of university students of different majors (mean years of education: 16.73 years, *SD* = 3.04 years), who either received a financial reward or course credits for their participation. Health status was self-reported by the participants as part of the demographic questionnaire. Individuals who reported current or past neurological or psychological problems were not admitted to the study. The study protocol was approved by the local ethics committee of the Faculty of Psychology at Ruhr University Bochum (vote Nr. 165). All participants gave written informed consent and were treated following the Declaration of Helsinki.

Acquisition and Analysis of Behavioral Data

Behavioral data was acquired during four individual test sessions. Each session was designed as a group setting of up to six participants, seated at individual tables, in a quiet and well-lit room. The tests were administered according to their respective manuals. The following is a brief description of each test procedure used in our study. Please refer to the Supplementary Material for descriptive statistics (Supplementary Material S1) and intercorrelations (Supplementary Material S2) of all cognitive tests.

I-S-T 2000 R. The *Intelligenz-Struktur-Test 2000 R* (I-S-T 2000 R) is a well-established German intelligence test battery measuring multiple facets of general intelligence [28,29]. The test consists of a basic and an extension module. The basic module measures different aspects of fluid intelligence and contains 180 items assessing verbal, numerical and figural abilities as well as 23 items assessing verbal and figural memory. Verbal, numeric, and figural abilities are measured by three reasoning tasks that comprise 20 items each. For instance,

verbal intelligence is assed via items on sentences completion, where the participant is asked to complete a sentence with the correct word, or on analogies and commonalities. Numerical intelligence on the contrary comprises items on arithmetic problems, digit spans, and arithmetic operators assessing the mathematical abilities of the participant. Figural intelligence is assessed via items testing for the participant's ability to assemble figures mentally, to mentally rotate and match dices, and to solve matrix-reasoning tasks. The processing time for this section is about 90 minutes. Subsequently, the participants complete two memory tasks, one verbal (10 items) and one figural (13 items), where they must memorize a series of words, or pairs of figures, respectively. This takes about 10 minutes. The extension module measures general knowledge covering a total number of 84 items. The knowledge test covers verbal (26 questions), numerical (25 questions), and figural knowledge (22 questions) and takes about 40 minutes. Here, the participant's knowledge on various facets is assessed: Art/literature, geography/history, mathematics, science, and daily life. Most of the items of both modules are designed in multiple-choice form. The only exception are two sub-tests on numerical reasoning (calculations and number series). Here the participant has to directly fill-in the answer. The complete testing session takes about 2 h 30 min. The reliability estimates (Cronbach's α) for the sub-factettes of the basic module fall between .88 and .95, as well as .93 for the memory tasks and .96 for general mental ability. The extension module has a reliability of .93 (Cronbach's α). The recent norming sample consists of about 5800 individuals for the basic module and 661 individuals for the extension module. The age range in the norming sample is between 15 - 60 years and both sexes are represented equally.

ZVT. The *Zahlenverbindungstest* (ZVT) is a trail-making test used to assess processing speed in both children and adults [30]. After editing two sample matrices, the participant has to process a total of four matrices. Here, the individual has to connect circled numbers from 1 to 90 in ascending order. The numbers were positioned more or less randomly within the matrix. The instructor measures the processing time for each matrix. The total test value, reflecting the participant's processing speed, is then calculated as the arithmetic mean of all four matrices. The test takes about 10 minutes in total. The reliability between the individual matrices is above .86 and .95 for adults and six months retest-reliability is between .84 and .90. The recent norming sample consists of about 2109 individuals with an age range between 8 - 60 years and equal sex representation.

BOMAT advanced short. The *Bochumer Matrizentest* (BOMAT) is a non-verbal intelligence test which is widely used in neuroscientific research [31-33,3]. Its structure is similar to the well-established Raven's Advanced Progressive Matrices [4]. Within the framework of our study we carried out the advanced short version of the BOMAT, which is known to have a high discriminatory power in samples with generally high intellectual abilities, thus avoiding possible ceiling effects [33,32]. The test comprises two parallel forms with 29 matrix-reasoning items

each. Participants were assigned to one of the two forms, which they had to complete. The participants have a total of 45 minutes to process as many matrices as possible. The split-half reliability of the BOMAT is .89, Cronbach's α is .92, and reliability between the parallel-forms is .86. The recent norming sample consists of about 2100 individuals with an age range between 18 - 60 years and equal sex representation.

BOWIT. The *Bochumer Wissenstest* (BOWIT) is a German inventory to assess the subject's degree of general knowledge [5]. The inventory comprises two parallel forms with 154 items each. Both forms include eleven different facets of general knowledge: Arts/architecture, language/ literature, geography/logistics, philosophy/religion, history/archeology, economics/law, civics/ politics, as well as, biology/chemistry, nutrition/health, mathematics/physics, and technology/ electronics. Within one test form, each knowledge facet is represented by 14 multiple-choice items. In our study, all participants had to complete both test forms resulting in a total number of 308 items. The processing time for each test form is about 45 minutes. The knowledge facets assessed by the BOWIT are very similar to general knowledge inventories used in other studies [34-36]. The inventory fulfills all important quality criteria regarding different measures of reliability and validity. The inventory's manual specifies that split-half reliability is .96, Cronbach's α is .95, test-retest reliability is .96 and parallel-form reliability is .91. Convergent and discriminant validity are given for both test forms. The norming sample consists of about 2300 individuals (age range: 18 - 66 years) and has an equal sex representation.

DNA sampling and genotyping

For non-invasive sampling, exfoliated cells were brushed from the oral mucosa of the participants. DNA isolation was performed with QIAamp DNA mini Kit (Qiagen GmbH, Hilden, Germany). Genotyping was carried out using the Illumina Infinium Global Screening Array 1.0 with MDD and Psych content (Illumina, San Diego, CA, USA) at the Life & Brain facilities, Bonn, Germany. Filtering was performed with PLINK 1.9 [37,38] removing SNPs with a minor allele frequency of <0.01, deviating from Hardy-Weinberg equilibrium with a *p*-value of $<1*10^{-6}$, and missing data >0.02. Participants were excluded with >0.02 missingness, sex-mismatch, and heterozygosity rate > |0.2|. Filtering for relatedness and population structure was carried out on a SNP set of filtered for high quality (HWE $p > 0.02$, MAF > 0.2, missingness = 0), and LD pruning ($r^2 = 0.1$). In pairs of cryptically related subjects (pi hat > 0.2), one subject was excluded at random. Principal components to control for population stratification were generated, and outliers > |6SD| on one of the first 20 PC were excluded. The final data set consisted of 296 participants and 491,138 SNPs.

Polygenic scores

We created genome-wide polygenic scores for each participant using publicly available summary statistics for intelligence and educational attainment [25,24]. Polygenic scores were constructed as the weighted sums of each participant's trait-associated alleles across all SNPs using PRSice 2.1.6 [39]. In regard to the highly polygenic nature identified for EA and IQ and the observed range of highest prediction of the PRS in the original manuscripts [24,25], we applied a *p*-value threshold (PT) of 0.05 for the inclusion of SNPs in the calculation of EA and IQ-PRS. Additionally, we report the results for the PGS with the strongest association with the respective cognitive tests (and subtests) in our sample (best-fit PGS). That is, the *p*-value threshold (PT) for inclusion of SNPs was chosen empirically (for the range of PT $5*10^{-8}$ - .5 in steps of $5*10^{-5}$), so the resulting PGS explained the maximum amount of test score variance for the respective measure in our sample. Finally, we also investigated the predictive power of PGS including all available SNPs (non-fit PGS), that is, the *p*-value threshold for SNP inclusion equaled PT = 1.00. The predictive power of the PGS derived from the GWAS was measured by the 'incremental R^2' statistic [25]. The incremental R^2 reflects the increase in the determination coefficient (R^2) when the PGS is added to a regression model predicting the behavioral phenotype alongside a number of control variables (here: sex, age, and the first four principal components of population stratification). For all statistical analyses in PRSice, linear parametric methods were used. Testing was two-tailed with an α-level of $p < .05$. As we report a total of 54 regression analyses, our results were FDR corrected for 54 multiple comparisons using the Benjamini-Hochberg method [40].

Results

In the following section, we report incremental determination coefficients (incremental R^2) for the PGS with a PT of 0.05 and respective test scores. Additionally, we investigated the association between our cognitive test scores and, so-called best-fit EA-PGS and IQ-PGS. These best-fit PGS were estimated by using a function, which empirically determines a *p*-value threshold for SNP inclusion to maximally predict the performace in the respective cognitive test (see Supplementary Material S3 and S4 for intercorrelations between best-fit PGS). Finally, we explored the predictive power of non-fit PGS including all available SNPs (PT = 1.00) and cognitive test scores.

For PT of 0.05 EA-PGS were especially predictive of individual differences in general intelligence (incremental R^2 = 3.25%, $p < .001$), numerical intelligence (incremental R^2 = 2.79%, $p = .001$) and verbal intelligence (incremental R^2 = 2.55%, $p < .001$), measured with the IST-2000-R (Figure 1). Aspects of general knowledge assessed by IST-2000-R (incremental R^2 = 1.57%, $p = .004$) and BOWIT (incremental R^2 = 1.19%, $p = .016$) were moderately predicted by EA-PGS. Differences in language-free components of general intelligence, measured as figural intelligence (incremental R^2 = 1.23%, $p = .017$) and matrices (incremental R^2 = 0.50%, p =.111), were partially only weakly explained by EA-PSG (Figure 1). Interestingly, EA-PGS had a high predictive power for processing speed (incremental R^2 = 2.22%, $p = .001$). The least predictive power was detected for individual differences in short-term memory performance (incremental R^2 = 0.34%, $p = .178$).

Individual differences in IQ-PGS were especially predictive of individual differences in verbal (incremental R^2 = 2.39%, $p < .001$) and general intelligence (incremental R^2 = 1.78%, $p = .002$) measured with the IST-2000-R. The predictive power for differences in numerical intelligence was only slightly lower (Figure 1). Here, IQ-PGS had an incremental R^2 of 1.54% ($p = .010$). Regarding measures of general knowledge, IQ-PGS had an incremental R^2 of 1.27% for ability differences in the IST-2000-R general knowledge test ($p = .010$) and 1.23% for differences assessed with the BOWIT ($p = .014$). Differences in non-verbal aspects of intelligence were only weakly predicted. Our multiple regression analyses resulted in predicitive values of R^2 = 0.22% for figural intelligence ($p = .315$), R^2 = 0.19% for matrices ($p = .326$), and R^2 = 0.10% for processing speed ($p = .489$). Again, the predictive value of PGS-IQ for short-term memory was minute (incremental R^2 = 0.22%, $p = .278$).

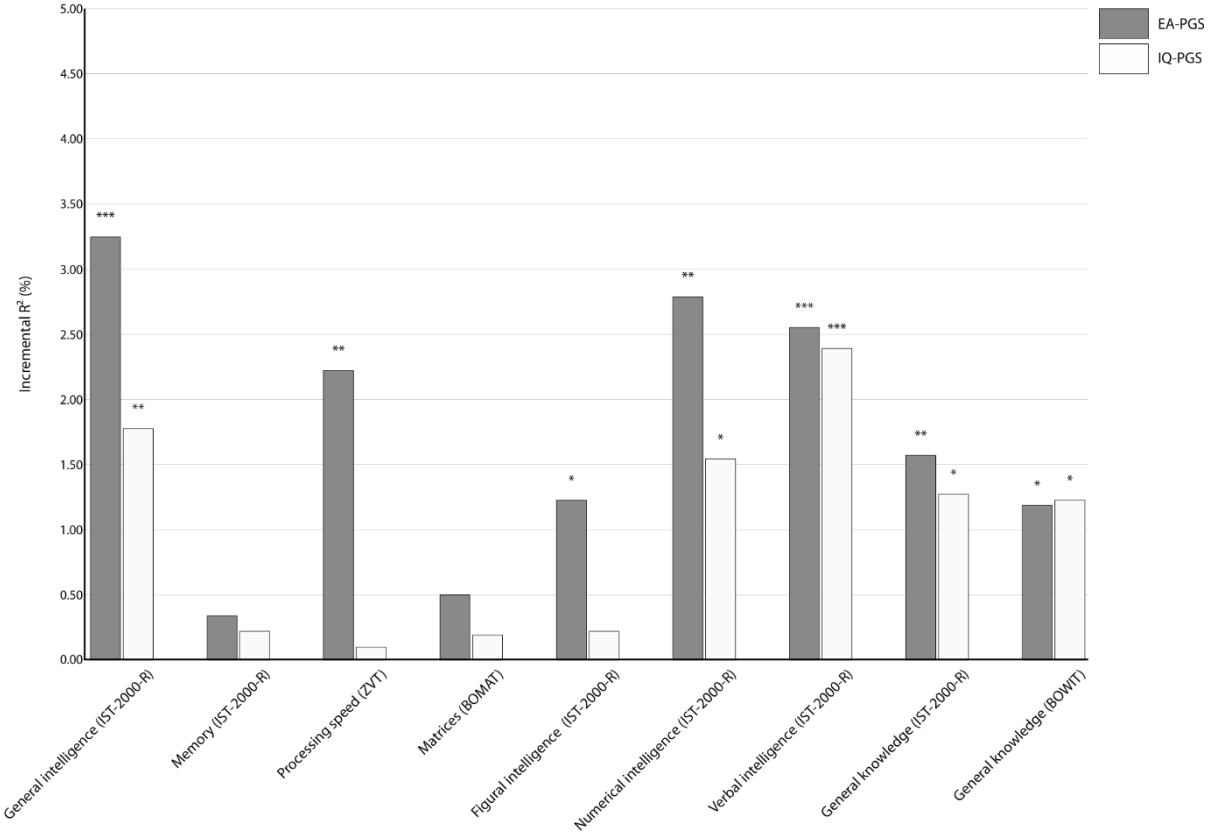

Figure 1. The figure depicts the incremental R^2 of the *P*-value threshold (PT) = .05 polygenic scores of educational attainment (EA-PGS) and intelligence (IQ-PGS) in percent. The incremental R^2 reflects the increase in the determination coefficient (R^2) when the EA-PGS or IQ-PGAS is added to a regression model predicting individual differences in the respective cognitive test. The association between PGS and phenotype was controlled for the effects of age, sex, population stratification, and multiple comparisons [40]. * adjusted $p \leq .05$, ** adjusted $p \leq .01$, *** adjusted $p \leq .001$.

Apart from the *p*-value threshold (PT) of 0.05, we also investigated the predictive power of PGS with the strongest association with the respective cognitive tests (and subtests) in our sample, so-called best-fit PGS.

Here best-fit EA-PGS were especially predictive of individual differences in verbal intelligence (incremental R^2 = 4.17%, $p < .001$), general intelligence (incremental R^2 = 4.03%, $p < .001$) and numerical intelligence (incremental R^2 = 3.92%, $p <.001$) measured with the IST-2000-R (Figure 2). Aspects of general knowledge assessed by BOWIT (incremental R^2 = 2.15%, p = .001) and IST-2000-R (incremental R^2 = 2.04%, p = .001) were moderately predicted by EA-PGS. In contrast, differences in language-free components of general intelligence, measured as figural intelligence (incremental R^2 = 1.78%, p = .004) and matrices (incremental R^2 = 1.13%, p =.017), were only weakly explained by EA-PSG (Figure 2). Interestingly, EA-PGS had a higher predictive power for processing speed (incremental R^2 = 2.61%, $p < .001$). The least predictive power was detected for individual differences in short-term memory performance (incremental R^2 = 1.11%, p = .014).

Individual differences in IQ-PGS were especially predictive of individual differences in general intelligence measured with the IST-2000-R (incremental R^2 = 3.18%, $p < .001$). The predictive

power for verbal intelligence and general knowledge was only slightly lower (Figure 2). Here, IQ-PGS had an incremental R^2 of 3.16% for verbal intelligence ($p < .001$) and 2.38% for differences in numerical intelligence ($p = .001$). Regarding measures of general knowledge, IQ-PGS had an incremental R^2 of 2.32 % for ability differences assessed with the BOWIT ($p = .001$) and 2.28% for differences in IST-2000-R general knowledge test ($p = .001$). Moreover, 1.80% of the difference in figural intelligence ($p = .004$) and 1.27% of the performance difference in the BOMAT ($p = .011$) were additionally explained by including IQ-PGS into the respective regression model. As depicted in Figure 2, IQ-PGS had only weak predictive power for processing speed (incremental $R^2 = 0.55\%$, $p = .100$) and short-term memory (incremental $R^2 = 0.30\%$, $p = .151$).

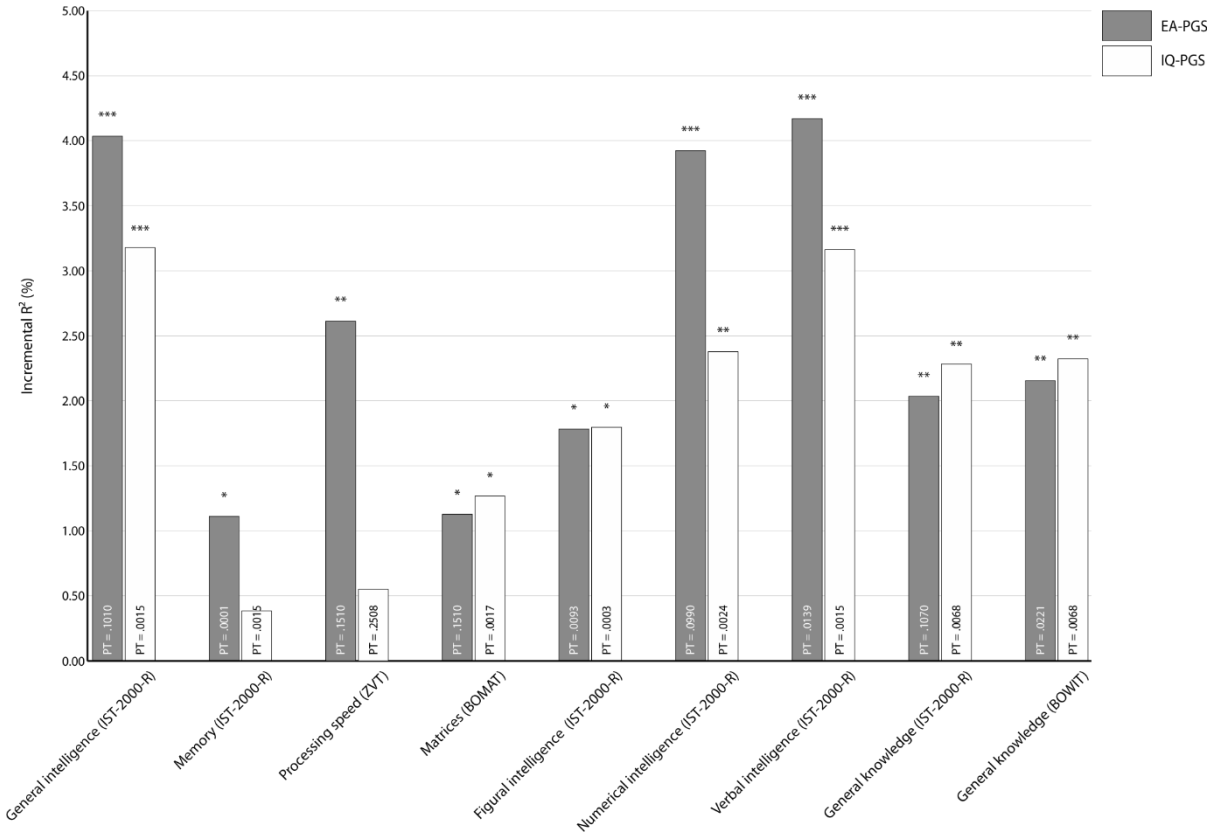

Figure 2. The figure depicts the incremental R^2 of the best-fit polygenic scores of educational attainment (EA-PGS) and intelligence (IQ-PGS) in percent. The *p*-value thresholds (PT) that determined the inclusion of SNPs into the respective PGS are displayed in the respective bar. The incremental R^2 reflects the increase in the determination coefficient (R^2) when the EA-PGS or IQ-PGAS is added to a regression model predicting individual differences in the respective cognitive test. The association between PGS and phenotype was controlled for the effects of age, sex, population stratification, and multiple comparisons [40]. * adjusted $p \leq .05$, ** adjusted $p \leq .01$, *** adjusted $p \leq .001$.

Finally, we investigated the predictive power of PGS summarizing the effects of all SNPs (PT = 1.00, non-fit PGS, Figure 3). Here, again the increment in the determination coefficient caused by EA-PGS was especially high for numerical (incremental $R^2 = 2.10\%$, $p = .003$), general (incremental $R^2 = 2.08\%$, $p = .001$), and verbal intelligence (incremental $R^2 = 1.75\%$,

p =.001) measured with the IST-2000 R. Moreover, EA-PGS increases the determination of the differences in general knowledge measured via IST-2000-R by 1.22% (both *p* =.012) and by 0.68% (*p* =.067) for knowledge differences assessed by BOWIT. While differences in processing speed were predicted with an incremental R^2 of 1.21% (*p* = .014), EA-PGS had the lowest incremental effect on predicting differences in figural intelligence (incremental R^2 = 0.50%, *p* =.130), matrices (incremental R^2 =0.60%, *p* =.082) and short-term memory (incremental R^2 =0.06%, *p* =.559).

A similar picture emerges for the predictive power of IQ-PGS. The highest predictive power occurred for measures of numerical intelligence (incremental R^2 = 1.94%, *p* =.004), general intelligence (incremental R^2 = 1.52%, *p* =.005), verbal intelligence (incremental R^2 = 1.27%, *p* =.006), and general knowledge (IST-2000-R: 0.79%, *p* =.042; BOWIT: 0.68%, *p* =.068). As shown in Figure 3, measures of non-verbal intelligence and sort-term memory were only poorly predicted by IQ-PGS (processing speed: R^2 = 0.27%, *p* =.292; matrices: R^2 = 0.18%, *p* =.342; figural intelligence: R^2 = 0.16%, *p* =.387; short-term memory: R^2 = 0.05%, *p* =.607).

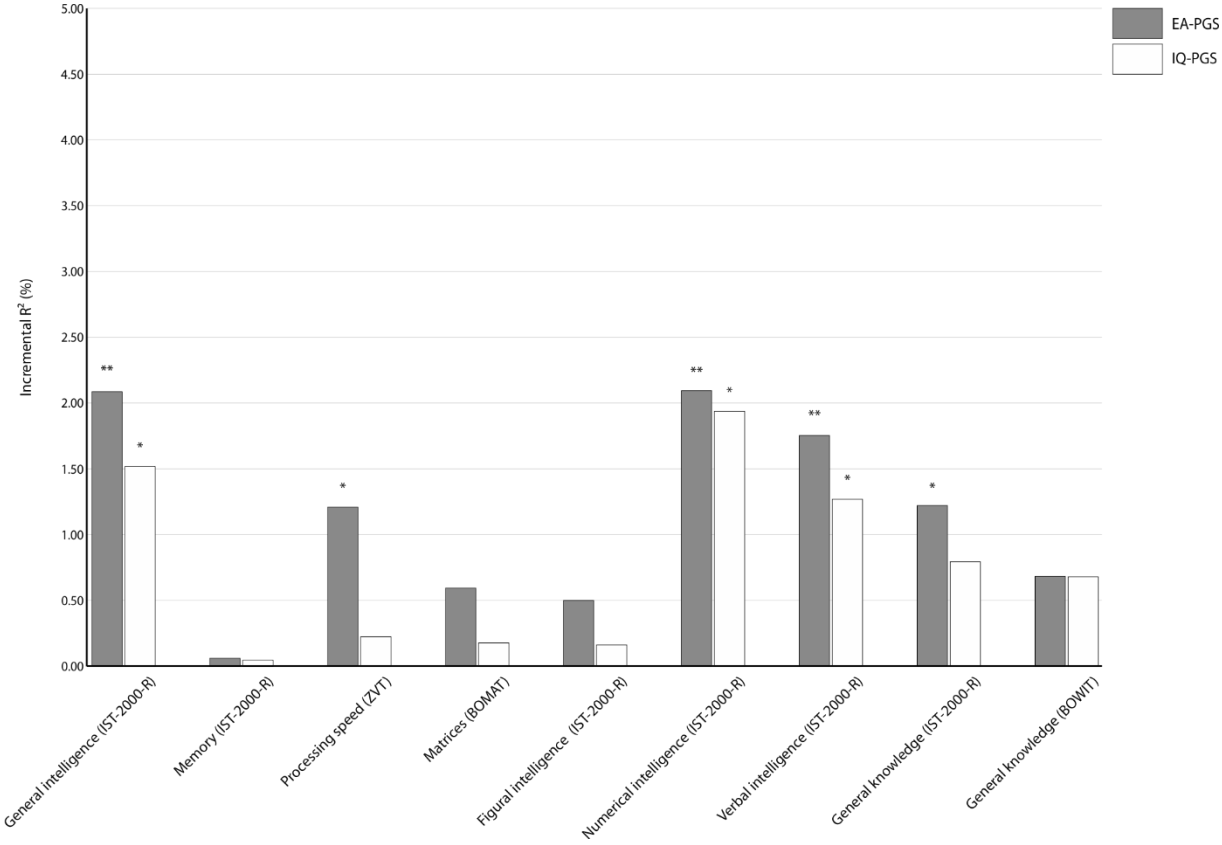

Figure 3. The figure depicts the incremental R^2 of the non-fit polygenic scores of educational attainment (EA-PGS) and intelligence (IQ-PGS) in percent. P-value thresholds (PT) = 1. The incremental R^2 reflects the increase in the determination coefficient (R^2) when the EA-PGS or IQ-PGAS is added to a regression model predicting individual differences in the respective cognitive test. The association between PGS and phenotype was controlled for the effects of age, sex, population stratification, and multiple comparisons [40]S. * adjusted *p* ≤ .05, *** adjusted *p* ≤ .01, **** adjusted *p* ≤ .001.

Overall, for the associations between cognitive measures and PT = 0.05, PT = 1.00 or best-fit PGS, a comparable pattern emerged, albeit the magnitude of explained variance deviated. Here, the predictive power of the PT = 0.05 and PT = 1.00 was lower than for the best-fit PGS. Moreover, the amount of explained variance was, in most cases, lower for IQ-PGS than for EA-PGS (Figure 1, Figure 2, and Figure 3).

Discussion

Polygenic scores for educational attainment and intelligence are increasingly used to investigate associations between genetic disposition for cognitive abilities and different life outcomes [21-23,25,24]. It is, however, currently not known which aspects of general intelligence are reflected to what extent by available PGS, derived from large GWAS. Here, we show that EA-PGS and IQ-PGS do not predict every form of cognitive ability equally well (Figure 1). Specifically, EA-PGS had a high predictive power for interindividual differences in general, verbal and numerical intelligence. In contrast, short-term memory was only weakly associated. A similar pattern was evident for IQ-PGS. Again, the PGS had the highest predictive power for individual differences in verbal and general intelligence, while non-verbal and memory-related facets of intelligence were only poorly predicted.

A potential cause for the deviations in the predictive power of the PGS is the design of the GWAS that were used to discover the PGS in the first place. In typical GWAS, sample sizes tend to be very large, which usually comes at the cost of light phenotyping. It follows that PGS, which summarize trait-associated effect sizes of single SNPs, will reflect the genetic basis of the measured phenotype. In the case of educational attainment, this was defined as the years of schooling an individual completed [25]. Because verbal and nummerical intelligence reflect culturally acquired abilities [2], it is not surprising, that PGS based on individual differences in years of education are primarily associated with individual differences in these aspects of intelligence, and less so with differences in non-verbal intelligence and short term memory.

Interestingly, the same applies to the IQ-PGS. These scores were based on a GWAS meta-analysis, that was mainly driven by the UK Biobank subsample. With almost 200,000 participants, the UK Biobank sample contributed about two-thirds of the meta-analysis sample [25]. In the UK Biobank, cognitive abilities were measured as the number of correct answers to a total of 13 questions assessing both verbal and mathematical intelligence. Although the respective questions are considered to measure intelligence in the form of verbal and numerical abilities, the number of correct answers seem to be dependent on culturally acquired knowledge rather than on deductive and inductive reasoning [41]. This becomes clear when taking a closer look at individual items. For example, arithmetic capability is measured with the following item: *"If David is twenty-one and Owen is nineteen and Daniel is nine years younger than David, what is half their age combined?" [41]*. As this is a classic word problem, it is not language-free and, therefore, clearly affected by cultural expertise. Again, the selection of test procedures used to determine individual differences in cognitive ability in the discovery sample appears to have a distorting influence on the predictive power of the PGS in other target samples. This is particularly important when one wants to investigate the functional pathways between genotype and phenotype. Researchers who aim to bridge the gap between genetic

and neuronal correlates of cognitive abilities should be aware of this issue as previous studies have shown that distinct aspects of intelligence have distinct neural correlates [33].

Although our study provides important insights into the predictive power of commonly used PGS, the following limitations need to be considered. First, our sample was mainly composed of university students with a restricted age range. Therefore, independent replication of our results in diverse samples is highly desireable [42]. Second, for the selection of a suitable *p*-value threshold for SNP inclusion, we were guided by the original publications of Savage et al. [24] and Lee et al. [25]. However, there is no clear specification as to which *p*-value threshold achieves the best overall predictive power regarding cognitive ability differences. Therefore, we additionally used a best-fit approach, which yielded different *p*-value thresholds for SNP inclusion for the different cognitive tests (Figure 2). Here, *p*-value thresholds for the various PGS ranged from PT = .0001 to PT = .1510 but were highly intercorrelated (see Supplementary Material S3 and S4) so that the different best-fit scores are highly comparable regarding their SNP composition. As using a best-fit approach can lead to an overestimation of the observed explained variance [39], we also report the associations of non-fit PGS with a *p*-value threshold of PT = 1.00 (Figure 3). With this approach, PGS that take all available SNPs for either intelligence and educational attainment into account were tested for associations with the different cognitive tests. Although PT = .05 and PT = 1.00 have a consistently lower predictive power than the best-fit PGS, they exhibit an comparable pattern of results. This confirms that both the polygenic scores of educational attainment and intelligence derived from large GWAS [25,24] reflect some aspects of intelligence, e.g., verbal and numerical intelligence, more accurately than others, like short-term memory.

In conclusion, our study is the first to investigate the association between polygenic scores of educational attainment [25] and intelligence [24] and different aspects of general intelligence in healthy adults. The study shows how large scale light-phenotyping GWAS studies with a strong statistical power to identify associated variants, and smaller comprehensive deep-phenotyping approaches with a fine-grained assessment of the phenotype of interest can complement each other. We demonstrate that the discovery GWAS for EA and IQ do not reflect every form of cognitive ability equally well. Realizing that the way of phenotyping in large GWAS affects the predictive power of the resulting PGS [see also 43] is essential for all future studies planning to use PGS to unravel the genetic correlates of cognitive abilities in their samples.

Funding

Funding for the research was provided by the Deutsche Forschungsgemeinschaft (DFG) grant number GU 227/16-1, GE 2777/2-1, and SFB 1280 project A03 and F02 (project number: 316803389), as well as the Mercur Foundation grant number An-2015-0044. FS acknowledges support by the German Federal Ministry of Education and Research (BMBF) through the ERA-NET NEURON, "SynSchiz - Linking synaptic dysfunction to disease mechanisms in schizophrenia - a multilevel investigation" (01EW1810) grant.

Conflicts of Interest

The authors cite no conflicts of interest and there are no competing financial interests with this work.

Ethics Approval

The study protocol was approved by the local ethics committee of the Faculty of Psychology at Ruhr University Bochum (vote Nr. 165).

Consent to Participate

All participants gave written informed consent and were treated following the Declaration of Helsinki.

Consent for publication

All participants gave written informed consent that their data might be part of scientific publications. In case of publication, it is ensured that the published data cannot be associated with individual participants at any time.

Availability of data and material

Since the data included in this study are part of an ongoing research project, they are not made openly accessible yet. However, if the manuscript is accepted for publication in Molecular Neurobiology, all necessary data will be made publicly available via an OSF link.

All cognitive ability socres were assessed via sandardized and published test instruments. Publication of the corresponding material is therefore not permitted due to copyright regulations.

Code availability

Analysis was applied with the freely available software PRSice 2.1.6. The specific code that was used for the analyses in this paper is available from the corresponding author upon reasonable request.

Author Contributions

E.G., R.K. and S.O conceived the project and supervised the experiments. E.G., C.S., L.A., H.P.N, O.G., R.K, and S.O. designed the project. L.A., H.P.N., F.S., and R.K. planned and performed genetic experiments. C.S. and C.F. collected data. E.G., C.S., L.A., M.C.V., C.F., R.K., and S.O. analyzed the data. E.G., C.S., F.S., R.K. and S.O. wrote the paper. All authors discussed the results and edited the manuscript.

E.G. Erhan Genç

C.S. Caroline Schlüter

C.F. Christoph Fraenz

L.A. Larissa Arning

H.P.N Huu Phuc Nguyen

M.C.V Manuel C. Voelkle

F.S. Fabian Streit

O.G. Onur Güntürkün

R.K. Robert Kumsta

S.O. Sebastian Ocklenburg

Acknowledgements

The authors would like to thank all research assistants for their support during the behavioral measurements.

References

1. Neisser U, Boodoo G, Bouchard Jr TJ, Boykin AW, Brody N, Ceci SJ, Halpern DF, Loehlin JC, Perloff R, Sternberg RJ (1996) Intelligence: knowns and unknowns. Am Psychol 51 (2):77. doi:Doi 10.1037/0003-066x.51.2.77

2. Flanagan DP, Dixon SG (2013) The Cattell-Horn-Carroll Theory of Cognitive Abilities. Encyclopedia of Special Education: A Reference for the Education of Children, Adolescents, and Adults with Disabilities and Other Exceptional Individuals

3. Hossiep R, Hasella M, Turck D (2001) BOMAT-advanced-short version: Bochumer Matrizentest. Hogrefe, Göttingen, Germany

4. Raven JC (1990) Court, J.H., & Raven, J. (1990). Coloured progressive matrices. Manual for Raven's Progressive Matrices and Vocabulary Scales

5. Hossiep R, Schulte M (2008) BOWIT: Bochumer Wissenstest. Hogrefe,

6. Roth B, Becker N, Romeyke S, Schafer S, Domnick F, Spinath FM (2015) Intelligence and school grades: A meta-analysis. Intelligence 53:118-137. doi:http://dx.doi.org/10.1016/j.intell.2015.09.002

7. Spengler M, Brunner M, Damian RI, Lüdtke O, Martin R, Roberts BW (2015) Student characteristics and behaviors at age 12 predict occupational success 40 years later over and above childhood IQ and parental socioeconomic status. Developmental psychology 51 (9):1329. doi:10.1037/dev0000025

8. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC (2004) The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. Journal of personality and social psychology 86 (1):130

9. Gottfredson LS, Deary IJ (2004) Intelligence predicts health and longevity, but why? Current Directions in Psychological Science 13 (1):1-4

10. Wraw C, Deary IJ, Der G, Gale CR (2016) Intelligence in youth and mental health at age 50. Intelligence 58:69-79

11. Knopik VS, Neiderhiser JM, DeFries JC, Plomin R (2016) Behavioral genetics. 7 edn. Macmillan Higher Education/W.H.Freeman & Co Ltd., London, UK

12. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2016) Top 10 replicated findings from behavioral genetics. Perspectives on psychological science 11 (1):3-23

13. Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Earth and Environmental Science Transactions of the Royal Society of Edinburgh 52 (2):399-433

14. Plomin R, von Stumm S (2018) The new genetics of intelligence. Nature Reviews Genetics 19 (3):148

15. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM (2014) Research Review: Polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry 55 (10):1068-1087. doi:10.1111/jcpp.12295

16. Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S, Hofer E, Ibrahim-Verbaas CA, Kirin M, Lahti J (2015) Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N= 53 949). Mol Psychiatr 20 (2):183-192

17. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JR, Krapohl E, Taskesen E, Hammerschlag AR, Okbay A, Zabaneh D (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nature Genetics 49 (7):1107

18. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra H-J, Shakhbazov K, Abdellaoui A, Agrawal A (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. science 340 (6139):1467-1471

19. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen G-B, Emilsson V, Meddens SFW (2016) Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533 (7604):539

20. Okbay A, Wedow R, Kong E, Turley P, Lee J, Zacher M, Thom K, Nguyen AT, Maghzian O, Linner RK (2017) GWAS of educational attainment: phase 3-main results. Behavior Genetics 47 (6):699-700

21. Belsky DW, Moffitt TE, Corcoran DL, Domingue B, Harrington H, Hogan S, Houts R, Ramrakha S, Sugden K, Williams BS, Poulton R, Caspi A (2016) The Genetics of Success:How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. Psychological science 27 (7):957-972. doi:10.1177/0956797616643070

22. Jansen PR, Polderman TJC, Bolhuis K, van der Ende J, Jaddoe VWV, Verhulst FC, White T, Posthuma D, Tiemeier H (2018) Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. J Child Psychol Psychiatry 59 (1):39-47. doi:10.1111/jcpp.12759

23. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, Zabaneh D, Breen G, O'Reilly PF, Plomin R (2018) Multi-polygenic score approach to trait prediction. Mol Psychiatr 23 (5):1368-1374. doi:10.1038/mp.2017.163

24. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, Nagel M, Awasthi S, Barr PB, Coleman JR (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature genetics 50 (7):912

25. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Linnér RK (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics 50 (8):1112

26. Cohen J (1992) A power primer. Psychological Bulletin 112 (1):155-159. doi:http://dx.doi.org/10.1037/0033-2909.112.1.155

27. Faul F, Erdfelder E, Buchner A, Lang A-G (2009) Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 41 (4):1149 - 1160. doi:https://doi.org/10.3758/BRM.41.4.1149

28. Liepmann D, Beauducel A, Brocke B, Amthauer R (2007) Intelligenz-Struktur-Test 2000 R (IST 2000 R). Manual (2. erweiterte und überarbeitete Aufl.). Göttingen: Hogrefe

29. Beauducel A, Brocke B, Liepmann D (2001) Perspectives on fluid and crystallized intelligence: facets for verbal, numerical, and figural intelligence. Pers Indiv Differ 30 (6):977-994. doi:Doi 10.1016/S0191-8869(00)00087-8

30. Oswald WD, Roth E (1987) Der Zahlen-Verbindungs-Test (ZVT). Hogrefe Verlag fuer Psychologie,

31. Oelhafen S, Nikolaidis A, Padovani T, Blaser D, Koenig T, Perrig WJ (2013) Increased parietal activity after training of interference control. Neuropsychologia 51 (13):2781-2790

32. Genç E, Fraenz C, Schlüter C, Friedrich P, Hossiep R, Voelkle MC, Ling JM, Güntürkün O, Jung RE (2018) Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. Nature communications 9 (1):1905. doi:https://doi.org/10.1038/s41467-018-04268-8

33. Genç E, Fraenz C, Schlüter C, Friedrich P, Voelkle MC, Hossiep R, Güntürkün O (2019) The Neural Architecture of General Knowledge. European Journal of Personality 33 (5):589-605

34. Ackerman PL, Beier ME, Bowen KR (2002) What we really know about our abilities and our knowledge. Pers Indiv Differ 33 (4):587-605. doi:10.1016/s0191-8869(01)00174-x

35. Lynn R, Ivanec D, Zarevski P (2009) Sex Differences in General Knowledge Domains. Collegium Antropol 33 (2):515-520

36. Bratko D, Butkovic A, Chamorro-Premuzic T (2010) The genetics of general knowledge: A twin study from Croatia. Pers Indiv Differ 48 (4):403-407. doi:10.1016/j.paid.2009.11.006

37. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4 (1). doi:10.1186/s13742-015-0047-8

38. Purcell S, Chang C PLINK. www.cog-genomics.org/plink/1.9/.

39. Choi SW, O'Reilly PF (2019) PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience 8 (7). doi:10.1093/gigascience/giz082

40. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57 (1):289-300

41. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine 12 (3):e1001779

42. Hanel PH, Vione KC (2016) Do student samples provide an accurate estimate of the general public? PloS one 11 (12):e0168354

43. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, Clarke T-K, Forstner AJ, Grabe HJ, Hamilton SP, Levinson DF, Lewis CM, Lewis G, Martin NG, Milaneschi Y, Mors O, Müller-Myhsok B, Penninx BWJH, Perlis RH, Pistis G, Potash JB, Preisig M, Shi J, Smoller JW, Streit F, Tiemeier H, Uher R, Van der Auwera S, Viktorin A, Weissman MM, Kendler KS, Flint J, Consortium MDDWGotPG (2020) Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nature Genetics 52 (4):437-447. doi:10.1038/s41588-020-0594-5