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

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

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

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

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

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

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

°: These authors contributed equally

*: Corresponding Author and Lead Contact

Dr. Erhan Genç

Department of Biopsychology, Institute of Cognitive Neuroscience

Faculty of Psychology, Ruhr University Bochum

Universitätsstraße 150, 44801 Bochum, Germany

Phone: +49 234 32 24323

Fax: +49 234 32 14377

email: erhan.genc@rub.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 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^2 for crystallized intelligence (4.17%; 3.16%) and the weakest for short-term memory (1.11%; 0.30%). 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, general intelligence

Introduction

Differences in cognitive abilities are of great importance not only for psychological research but also for our everyday lives. 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 (Neisser et al., 1996). Various tests have been designed to measure the cognitive abilities of a person by assessing different aspects of general intelligence. Here, the two main components of general intelligence, fluid, and crystallized intelligence, are particularly important (Flanagan & Dixon, 2013). While inductive and deductive reasoning abilities are regarded as fluid intelligence, crystallized intelligence primarily comprises the amount of acquired declarative knowledge (Flanagan & Dixon, 2013). Whereas most test procedures cover both aspects of general intelligence, there are also tests that focus on either fluid or crystallized intelligence. Matrix reasoning tests are typically used to assess fluid intelligence (Hossiep, Hasella, & Turck, 2001; Raven, 1990), while crystallized intelligence comprises tests measuring aspects of declarative knowledge. This is stored long-term memory of static information like rules, relationships, abstract concepts, and of course, general knowledge (Flanagan & Dixon, 2013; Hossiep & Schulte, 2008).

Decades of intelligence research have shown that general intelligence is one of the best predictors of important life outcomes, including educational and occupational success (Roth et al., 2015; Spengler et al., 2015) as well as mental and physical health (Deary, Whiteman, Starr, Whalley, & Fox, 2004; Gottfredson & Deary, 2004; Wraw, Deary, Der, & Gale, 2016). 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 Knopik, Neiderhiser, DeFries, & Plomin, 2016; Plomin, DeFries, Knopik, & Neiderhiser, 2016). 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 (Fisher, 1918), 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) (Plomin & von Stumm, 2018).

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 (Wray et al., 2014). For instance, PGS for intelligence (IQ-PGS) (Davies et al., 2015; Sniekers et al., 2017) and educational attainment (EA-PGS) (Okbay et al., 2016; Okbay et al., 2017; Rietveld et al., 2013), 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 (Belsky et al., 2016; Jansen et al., 2018; Krapohl et al., 2018). Although IQ-PGS and EA-PGS explain a considerable amount of variance in intelligence (which is thought to increase even further with larger GWAS) (Plomin & von Stumm, 2018), 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 (Savage et al., 2018), or even more crudely, years of education (Lee et al., 2018). The question thus arises, which of the various aspects of general intelligence (Flanagan & Dixon, 2013; Neisser et al., 1996), 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 (Lee et al., 2018) and IQ-PGS (Savage et al., 2018) 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. (2018) and Lee et al. (2018). Both meta-analyses reported effect sizes (Cohen, 1992) of R^2 = .052 (IQ-PSG) (Savage et al., 2018), and R^2 =.097 (EA-PGS) (Lee et al., 2018), respectively. G-power (Faul, Erdfelder, Buchner, & Lang, 2009) 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 a more detailed description of all tests (Supplementary Material S1).

IST-2000-R. The Intelligenz-Struktur-Test 2000 R (IST-2000-R) is a well-established German intelligence test battery measuring multiple facets of general intelligence (Beauducel, Brocke, & Liepmann, 2001; Liepmann, Beauducel, Brocke, & Amthauer, 2007). The test consists of a basic and an extension module. The basic module contains 203 items assessing verbal, numeric, and figural reasoning as well as short term memory. The general intelligence score results from the total number of correct answers, while scores for verbal, numerical, and figural

intelligence, as well as short term memory, reflect correct responses within the respective test section. The processing time for the basic module is about 100 minutes. The extension module measures general knowledge with a total number of 84 items. Again, the test score reflects the amount of correctly answered items. The processing time of the extension module is 45 minutes.

ZVT. The *Zahlenverbindungstest* (ZVT) is a trail-making test commonly used to assess processing speed in both children and adults (Oswald & Roth, 1987). After editing two sample matrices, the participant has to process a total of four matrices. Here, the individual must 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.

BOMAT advanced short. The *Bochumer Matrizentest* (BOMAT) is a non-verbal intelligence test, similar to the well-established Raven's Advanced Progressive Matrices (Hossiep et al., 2001; Raven, 1990). 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 (Genç et al., 2018; Genç et al., 2019). The test comprises two parallel forms (A and B) with 29 matrix-reasoning items each. Participants were assigned to one of the two forms. Participants have a total of 45 minutes to process as many matrices as possible. The BOMAT test score reflects the number of successfully solved matrices.

BOWIT. The *Bochumer Wissenstest* (BOWIT) is a German inventory to assess a subject's degree of general knowledge (Hossiep & Schulte, 2008). The inventory comprises two parallel forms (A and B) with 154 items each. Both forms include different facets of general knowledge. In our study, 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 overall test score results from the total number of correct answers.

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 (Chang et al., 2015; Purcell & Chang) 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 (Lee et al., 2018; Savage et al., 2018). Polygenic scores were constructed as the weighted sums of each participant's trait-associated alleles across all SNPs using PRSice 2.1.6 (Choi & O'Reilly, 2019). The function of PRSice is to determine the threshold of each 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. Additionally, we report the association results for the non-fit PGS including all available SNPs, 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 (Lee et al., 2018). 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 36 regression analyses, our results were FDR corrected for 36 multiple comparisons using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995).

Results

In the following section, we report incremental determination coefficients (incremental R^2) for the best-fit PGS and respective test scores (see Methods and Supplementary Material for details). Based on these test scores EA-PGS and IQ-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 (best-fit PGS). Additionally, we investigated the association between non-fit PGS (including all available SNPs) and cognitive test scores. Information on the *p*-value thresholds for the best-fit PGS are depicted in Figure 1. For the intercorrelation between scores based on different inclusion criteria please refer to the Supplementary Material (Supplementary Material S2 and S3).

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 1). 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 1). 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 crystallized facets of intelligence assessed via measures of verbal intelligence and general knowledge was only slightly lower (Figure 1). 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 1, 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%, $p = .151$).

Finally, the overall amount of explained variance was, in most cases, lower for IQ-PGS than for EA-PGS (Figure 1). For the associations between cognitive measures and non-fit PGS,

using the most liberal threshold of PT = 1.00, a comparable pattern emerged, albeit with overall lower amounts of explained variance (Supplementary Material S4).

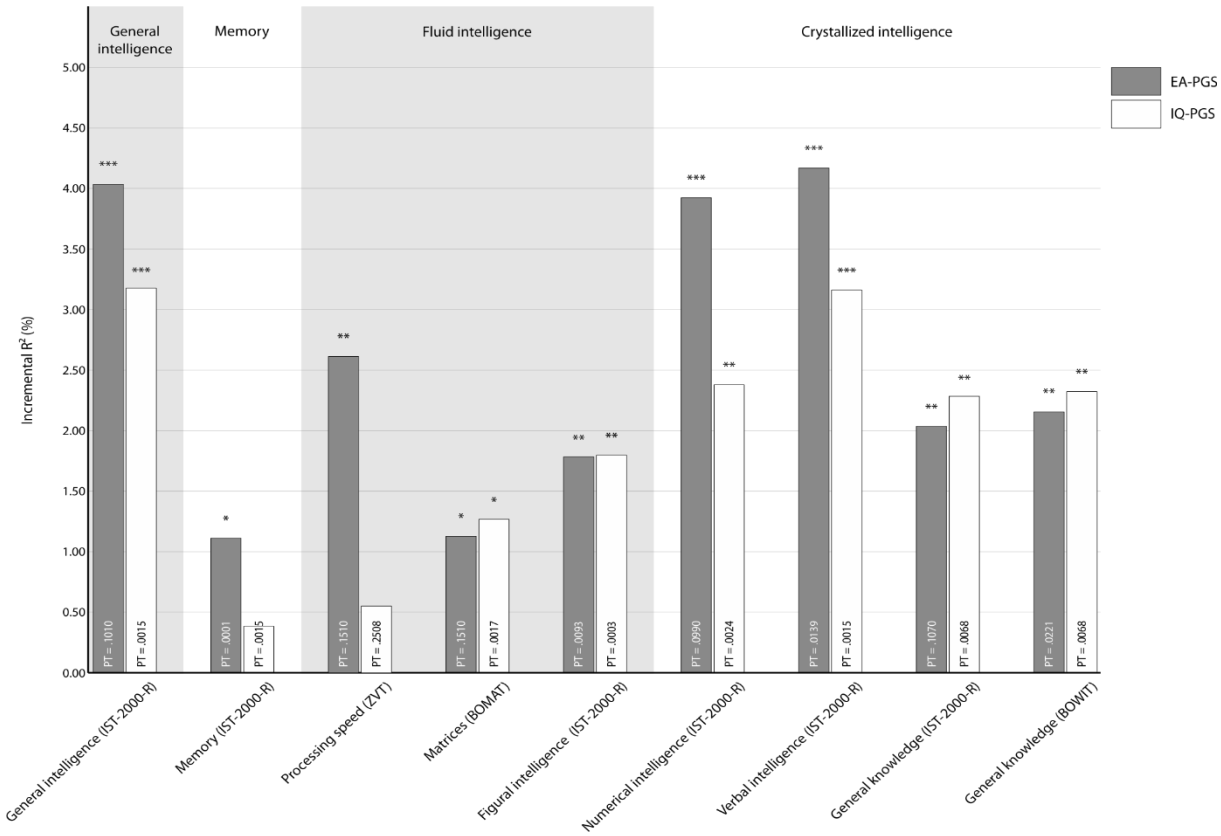

Figure 1. The figure depicts the incremental R^2 of the best-fit polygenic scores of educational attainment (EA-PGS) and intelligence (IQ-PGS) in percent. The *p*-value thresholds (PT) that determined the inclusion of SNPs into the respective PGS are displayed in the respective bar. The incremental R^2 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 (Benjamini & Hochberg, 1995). * adjusted $p \leq .05$, ** adjusted $p \leq .01$, *** adjusted $p \leq .001$.

Discussion

Polygenic scores for educational attainment and intelligence are increasingly used to investigate associations between genetic disposition for cognitive abilities and different life outcomes (Belsky et al., 2016; Jansen et al., 2018; Krapohl et al., 2018; Lee et al., 2018; Savage et al., 2018). 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. Specifically, EA-PGS had a high predictive power for interindividual differences in crystallized intelligence, namely, verbal and numerical intelligence, as well as general 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 aspects of crystallized intelligence, namely verbal intelligence, and general knowledge. In contrast, fluid 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 (Lee et al., 2018). Because crystallized intelligence reflects our culturally acquired abilities, such as verbal or numerical skills and declarative knowledge (Flanagan & Dixon, 2013), it is not surprising, that PGS based on individual differences in years of education are primarily associated with individual differences in crystallized intelligence, and less so with differences in fluid 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 (Lee et al., 2018). 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 fluid intelligence in the form of verbal and numerical abilities, the number of correct answers seems to dependent on culturally acquired knowledge rather than on deductive and inductive reasoning (Sudlow et al., 2015). 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?" (Sudlow et al., 2015)*. 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 fluid and crystallized intelligence have distinct neural correlates (Genç et al., 2019).

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 (Hanel & Vione, 2016). Second, as we aimed to identify the maximum amount of variance that could be explained by EA-PGS and IQ-PGS, respectively, we used a best-fit approach, which yielded different *p*-value thresholds for SNP inclusion for the different cognitive tests. *P*-value thresholds for the various PGS ranged from PT = .0001 to PT = .1510 but were highly intercorrelated (see Supplementary Material S2 and S3) so that the different best-fit scores are highly comparable regarding their SNP composition. Moreover, using the best-fit approach can lead to an overestimation of the observed explained variance (Choi & O'Reilly, 2019). To further address these issues, we report the associations of non-fit PGS with a *p*-value threshold of PT = 1.00 in Supplementary Material S4. With this approach, the exact same PGS (taking all available SNPs into account) for intelligence and educational attainment respectively were tested for associations with the different cognitive tests, making the results more comparable. Although the non-fit PGS have a consistently lower predictive power than the best-fit PGS, they exhibit an identical pattern of results (Supplementary Material S4). This confirms that both the best-fit and non-fit polygenic scores of educational attainment and intelligence derived from large GWAS (Lee et al., 2018; Savage et al., 2018) primarily reflect aspects of crystallized intelligence.

In conclusion, our study is the first to investigate the association between polygenic scores of educational attainment (Lee et al., 2018) and intelligence (Savage et al., 2018) 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. More specifically, both EA-PGS and IQ-PGS were especially predictive of individual differences in crystallized intelligence. Realizing that the way of phenotyping in large GWAS affects the predictive power of the resulting PGS (see also Cai et al., 2020) is essential for all future studies planning to use PGS to unravel the genetic correlates of cognitive abilities in their samples.

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

This work was supported 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. Finally, the authors would like to thank all research assistants for their support during the behavioral measurements.

References

Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences, 30*(6), 977-994.

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., et al. (2016). The Genetics of Success:How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. *Psychological Science, 27*(7), 957-972.

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.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., et al. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics, 52*(4), 437-447.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience, 4*(1).

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience, 8*(7).

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., et al. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N= 53 949). *Molecular psychiatry, 20*(2), 183-192.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (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.

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.

Fisher, R. A. (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.

Flanagan, D. P., & Dixon, S. G. (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*.

Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., et al. (2018). Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. *Nature Communications, 9*(1), 1905.

Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Voelkle, M. C., Hossiep, R., et al. (2019). The Neural Architecture of General Knowledge. *European Journal of Personality, 33*(5), 589-605.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13*(1), 1-4.

Hanel, P. H., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PLoS ONE, 11*(12), e0168354.

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

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

Jansen, P. R., Polderman, T. J. C., Bolhuis, K., van der Ende, J., Jaddoe, V. W. V., Verhulst, F. C., et al. (2018). Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *Journal of Child Psychology and Psychiatry, 59*(1), 39-47.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioral genetics* (7 ed.). London, UK: Macmillan Higher Education/W.H.Freeman & Co Ltd.

Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., von Stumm, S., Dale, P. S., et al. (2018). Multi-polygenic score approach to trait prediction. *Molecular Psychiatry, 23*(5), 1368-1374.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (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.

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*.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: knowns and unknowns. *American psychologist, 51*(2), 77.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539.

Okbay, A., Wedow, R., Kong, E., Turley, P., Lee, J., Zacher, M., et al. (2017). GWAS of educational attainment: phase 3-main results. *Behavior Genetics, 47*(6), 699-700.

Oswald, W. D., & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT)*: Hogrefe Verlag fuer Psychologie.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on psychological science, 11*(1), 3-23.

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

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

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

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science, 340*(6139), 1467-1471.

Roth, B., Becker, N., Romeyke, S., Schafer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53*, 118-137.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics, 50*(7), 912.

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R., Krapohl, E., et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics, 49*(7), 1107.

Spengler, M., Brunner, M., Damian, R. I., Lüdtke, O., Martin, R., & Roberts, B. W. (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.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (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.

Wraw, C., Deary, I. J., Der, G., & Gale, C. R. (2016). Intelligence in youth and mental health at age 50. *Intelligence, 58*, 69-79.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, 55*(10), 1068-1087.