Research Article

Evidence for Recent Polygenic Selection on Educational Attainment and Intelligence Inferred from GWAS Hits: A Replication of Previous Findings using Recent Data

Davide Piffer*

Ben Gurion University of the Negev, Israel

*Corresponding author: Davide Piffer, Ben Gurion University of the Negev, Israel. Tel: +972525811001; Email: pifferdavide@gmail.com

Abstract

Background: The genetic variants identified by three large genome-wide association studies (GWAS) of educational attainment and the largest intelligence GWAS were used to test a polygenic selection model.

Methods: Average frequencies of alleles with positive effect (polygenic scores or PS) were compared across populations (N=26) using data from 1000 Genomes. Factor analysis was used to extract a signal of polygenic selection.

Results: A polygenic selection factor of educational attainment GWAS hits is high among a handful of SNPs within genomic regions replicated across GWAS publications and it is highly correlated to the genetic intelligence factor (r= 0.96). These factors are both highly predictive of average population IQ (r=0.9), and are robust to tests of spatial autocorrelation. Several Monte Carlo simulations yielded highly significant p values. Furthermore, the polygenic selection model shows high replicability, with the EA and intelligence factor scores being virtually identical to those from an older study (r=0.96-0.99).

Conclusion: This study shows robust results after accounting for spatial autocorrelation and Monte Carlo simulation using random SNPs and shows robust reproducibility of results from a previous study.

Keywords: Educational Attainment; GWAS; IQ; Polygenes; Polygenic Selection

Introduction

Over the last decade, population geneticists have recognized that most traits are highly polygenic, and hence have moved away from the study of genetic evolution using the single-gene, Mendelian approach, towards models that examine many genes together (i.e. polygenic models).

Signals of polygenic selection can be identified by various methods, such as correlation of allele frequencies [1-4] and the regression of population average of trait values on Polygenic Scores (PS) [2,5-7], which have been successfully applied to human stature [5-7] and cognitive abilities [2]. This paper has several aims: to test the presence of a factor among GWAS hits and the predictive power of polygenic scores (average frequencies of GWAS alleles with positive effect), independently of spatial autocorrelation. A prediction is that the polygenic selection model provides a better fit to the data (i.e. average population IQ). A goal of this study is to replicate the effects found by Piffer in 2015 [2], with the evidence from new intelligence and educational attainment GWAS published to date. Piffer [2] factor analyzed educational attainment and intelligence GWAS hits and found a factor that was highly predictive of population IQ.

The factor analytic method is based on the assumption that polygenic selection acts as a latent variable which accounts for commonalities among several genetic variants scattered across the genome [1]. The model also includes an error term due to measurement error in the form of limited sample size, imperfect coverage or genetic drift, which all act to increase the noise.

Piffer [8] identified 9 genomic loci that were replicated across the three largest GWAS of educational ("EA") attainment

published to date [9-11]. The 9 loci contain GWAS significant alleles that were found to be in strong LD (r>0.8). One locus was replicated across three GWAS [9-11] and the same SNP (rs9320913) was found in two of them [9,11]. The population frequencies of the 9 pairs (one member belonging to each GWAS publication) of alleles were highly correlated (r=0.919), hence the SNPs published in [4] were used. Thus, this set of 9 SNPs was considered the best candidate for analysis of natural selection on educational attainment and related phenotypes (e.g. general cognitive ability or gca). In addition, the full sets of 74 and 162 SNPs (respectively, the new hits and those found after pooling together different datasets) from the latest GWAS of educational attainment [11] will be employed. Finally, the results of a recent GWAS of human intelligence [12] will be analyzed with the same method and compared to previous findings for educational attainment.

Average estimated population IQ will be used as the phenotype of interest and main dependent variable in the analyses. This choice can be justified by its privileged status in psychometric research and its robust genetic correlation (r= around 0.7) with educational performance [13] and attainment [14]. Moreover, the GWAS hits identified by the three educational GWAS also predict general cognitive ability in their samples [9-11]. A re-analysis of the Okbay et al. dataset revealed that the polygenic score also predicts general intelligence (3.6%) compared to 2% for the 2013 polygenic score [13]. If educational attainment and intelligence are genetically correlated also at the group level, a prediction is that there will be a population-level genetic correlation between educational attainment and intelligence GWAS polygenic or factor scores.

Materials and Methods

Rietveld et al. [9] produced 3 SNPs reaching GWAS significance for educational attainment. Davies et al. [10] reported 1115 SNPs reaching GWAS significance, of which 15 were independent signals for educational attainment. 942 SNPs were found on 1000 Genomes. Among the 15 independent signals, one (2:48696432_G_A) was missing.

Okbay et al. [11] reported 74 SNPs associated with years of education. 70 were found in 1000 Genomes (the other 4 variants were flagged because they had more than 3 different alleles). In a pooled meta-analysis, 162 SNPs were reported (161 were found in 1000 Genomes).

A set of nine GWAS hits that were in close linkage disequilibrium (linkage cut-off r>.8, 500kb linkage window) with 'hits' predicting educational attainment across two other large GWAS studies [9,10] was identified. Linkage was determined using the NIH LDLink program [14] with the 1000 Genomes Phase3 CEU population as a reference group. The 18 independent (i.e. unlinked) hits for general cognitive ability from the most recent intelligence GWAS were included in the analysis [12]. This study

also produced 9 SNPs that were also available in the educational attainment sample, and reached genome-wide significance [12]. This set of SNPs will be called "Int-EA SNPs".

Factor analysis was carried out using the Ordinary Least Squares method ("minres" option in R package "psych"). Monte Carlo simulations were performed using a random dataset, consisting of a large sample (N=13130) of matched random unlinked SNPs (downloaded from 1000 Genomes, phase 3). The empirical value p= (r+1)/(n+1) was calculated, where r is the number of runs whose Pearson's correlation coefficient (r x population IQ) was higher than the one found using the actual (GWAS-derived) polygenic score; n= total number of runs. The corrected formula was provided by Davison & Hinkley [15].

Matching was carried out using SNPSNAP [16], by feeding the 9 and the 18 intelligence SNPs and setting LD r^2 <0.1 (for EUR), with maximum allowed deviation for MAF= 5%. Unlinked SNPs were used in order to have a sample with independent observations. Fst distances were obtained from [2], calculated using Vcftools with 1000 Genomes, phase 3 data. Average population IQ estimates were obtained from [8]. Previously published scores were used also to guarantee that the values were not created ad hoc. SNP frequencies for ALFRED were downloaded from alfred. med.yale.edu. Fst distances for ALFRED/HGDP populations were obtained from [17], and after removal of the non-overlapping samples, 49 matching populations were retained. Analyses were run using R [18].

Results

Correlation Between Polygenic Scores and Population IQ

The polygenic score computed using the 9 SNPs was highly correlated (r=0.88) to an estimate [2] of average population IQ (Figure 1).
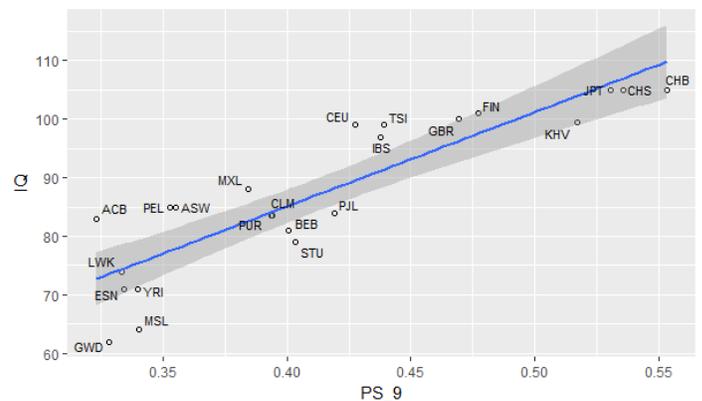

Figure 1: Correlation between population IQ and educational attainment polygenic score.

A Monte Carlo simulation was run using 818 PS computed from groups of 9 SNPs taken from the random dataset. The average correlation between population IQ and the random polygenic scores was 0.22 (N=818). The slightly positive correlation can be interpreted as an effect of spatial/phylogenetic autocorrelation [8] A Monte Carlo approach was used: the percentile corresponding to a correlation coefficient r=0.88 was found to be 99% (using the 818 random polygenic scores), implying that the result is highly significant the corrected (and more conservative) calculation of Monte Carlo p value, where p= r+1/n+1 (see Methods) was used, producing p= 0.011 (n= 819, r= 8).

The correlation between the 161 SNPs polygenic score and population IQ was high (r=0.854), corresponding to the 98th percentile in the simulation (using PS comprising 161 random SNPs each). The 161 SNPs and the 9 SNPs polygenic scores were strongly correlated (r= 0.949).

A polygenic score for the intelligence GWAS was computed, yielding only a moderate correlation with population IQ (r= 0.496), and a non-significant Monte Carlo p=0.43. The polygenic score had correlations of similar magnitude to the educational attainment polygenic scores, but had poor Monte Carlo significance (p=0.213). Since the population differences in polygenic scores were small, continent-level polygenic scores were computed to identify potential differences at a higher level (i.e. 1000 Genomes "superpopulations").

These only showed an advantage for East Asians, but were virtually identical for other superpopulations (EAS= 49%, EUR= 45.9%, AFR=47.3%, AMR= 45.6%, SAS= 44.2%). A polygenic score for the intelligence-educational attainment SNPs [12] was computed, yielding high correlations with population IQ (r= 0.876). Weighing by effect size produced almost identical results. Values for super-populations were the following: EAS= 62.98%, EUR= 56.95%, AFR= 50.71%, AMR=55.08%, SAS= 54.08%.

A Monte Carlo simulation for the correlation with population IQ was run, using SNP sets of the same size (9), yielding a highly significant p value= 0.008.

Factor Analysis of Allele Frequencies

Factor analysis was performed on the 9 quasi-replicated educational attainment SNP and on the 18 intelligence SNPs and on the 9 intelligence-EA SNPs, in order to extract a signal of polygenic selection [1,2]. Factor loadings are reported in (Tables 1a-c).

SNP	Loading
rs10191758	0
rs10236197	-0.382
rs11138902	0.71
rs12744310	-0.778
rs12928404	-0.329
rs13010010	0.929
rs16954078	-0.396
rs2251499	0.238
rs2490272	0.921
rs36093924	0.568
rs4728302	-0.713
rs6746731	-0.422
rs6779302	0.829
rs7646501	-0.977
rs9320913	0.84
rs66495454	0.726
rs113315451	-0.834
rs41352752	0.31

Table 1a: Factor loadings (Intelligence SNPs).

SNP	Loading
rs1008078	-0.801
rs11588857	0.831
rs12987662*	0.956
rs148734725	-0.492
rs11712056	0.606
rs62263923	0.863
rs13294439	0.91
rs12969294	0.459
rs9320913	0.717

Table 1b: Factor loadings (EA SNPs).

SNP	Loading
rs12928404	-0.358
rs12744310	-0.883
rs13010010	0.994
rs10191758	0.156
rs6801153	0.305
rs6779302	0.875
rs9320913	0.789
rs215601	0.475
rs4728302	0.510

Table 1c: Factor loadings (Intelligence-EA SNPs).

Average factor loadings were 0.046, 0.449 and 0.318 for the intelligence, the EA SNPs, and the intelligence-EA factors, respectively. Factor scores are reported in (Table 2).

Population	G factor	EA factor	Int-EA factor
ACB	-1.276	-1.351	-1.063
ASW	-0.961	-1.177	-0.997
BEB	-0.075	-0.209	-0.66
CDX	1.35	1.017	1.251
CEU	0.844	0.471	0.754
CHB	1.109	1.511	1.374
CHS	1.208	1.382	1.635
CLM	0.357	0.010	-0.113
ESN	-1.660	-1.453	-1.255
FIN	0.771	0.702	0.581
GBR	0.797	0.745	0.782
GIH	-0.049	0.271	-0.001
GWD	-1.358	-1.397	-1.186
IBS	0.631	0.350	0.476
ITU	-0.074	0.049	-0.212
JPT	0.878	1.342	1.321
KHV	1.267	1.346	1.925
LWK	-1.599	-1.488	-1.255
MSL	-1.444	-1.403	-1.165
MXL	0.215	0.056	-0.259
PEL	-0.060	0.050	-0.762
PJL	0.066	0.240	0.035
PUR	0.375	-0.004	-0.208
STU	-0.391	0.134	-0.432
TSI	0.764	0.248	0.677
YRI	-1.684	-1.443	-1.243

Table 2: Factor scores for educational attainment and intelligence.

Correlation Between Factors

The correlation between the G and EA factors was r= 0.96. A Monte Carlo simulation was run, carrying factor analysis over sets of 18 SNPs and correlating them to the EA 9 factor scores. The MC p value was 0.02. The Int-EA factor was similarly highly correlated to the other two factors (r= 0.93-0.94).

Controlling for Spatial Autocorrelation

The presence of spatial autocorrelation in a dataset means that the cases are not independent leading to an overestimation of degrees of freedom and, in the case of positive autocorrelation, an inflation in the correlation between two or more variables. The source of spatial autocorrelation in population genetics datasets is the similarity caused by admixture among neighboring populations, and the differences caused by random drift. Demonstrating that the alleles predict population-level differences in average phenotypic values above and beyond that predicted on the basis of migration, drift etc, provides evidence for a model of polygenic selection.

Fst distances were used to partial out spatial autocorrelation, following the method outlined in [1], similar to Mantel test [19]. The correlation between Fst distances and IQ distances was moderate (r=0.588), pointing out the presence of spatial autocorrelation. Multiple regression was performed with 9 SNPs and Fst as predictors and population IQ as dependent variable. Significant models (respectively for 9 and 161 SNPs) were obtained (Adjusted R-squared: 0.503, F-statistic: 128.5 on 2 and 250 DF, p< 2.2e-16), (Adjusted R-squared: 0.7251, F-statistic: 30.01 on 2 and 20 DF, p= 9.514e-07) (Table 3).

Variable	Beta	t	sig	VIF
PS 9 distances	0.524	9.024	<2e-16	1.713
Fst distances	0.250	4.305	2.4e-05	
PS 161 distances	0.456	7.063	1.61e-11	1.912
Fst dist	0.274	4.241	3.14e-05	
Fa EA dist	0.81	8.864	< 2e-16	4.22
Fst dist	-0.12	1.172	0.189	
Int factor dist	0.837	16.295	< 2e-16	3.42
Fst dist	-0.108	0.964	0.152	

*ACB= African Caribbeans in Barbados; ASW= Americans of African Ancestry in SW USA; BEB= Bengali from Bangladesh; CDX= Chinese Dai in Xishuangbanna, China; CEU= Utah Residents (CEPH) with Northern and Western European Ancestry; CHB= Han Chinese in Bejing, China; CHS= Southern Han Chinese; CLM= Colombians from Medellin, Colombia; ESN= Esan in Nigeria; FIN= Finnish in Finland; GBR= British in England and Scotland; GIH= Gujarati Indian from Houston, Texas; GWD= Gambian in Western Divisions in the Gambia; IBS= Iberian Population in Spain; ITU= Indian Telugu from the UK; JPT= Japanese in Tokyo, Japan; KHV= Kinh in Ho Chi Minh City, Vietnam; LWK= Luhya in Webuye, Kenya; MSL= Mende in Sierra Leone; MXL= Mexican Ancestry from Los Angeles USA; PEL= Peruvians from Lima, Peru; PJL= Punjabi from Lahore, Pakistan; PUR= Puerto Ricans from Puerto Rico; STU= Sri Lankan Tamil from the UK; TSI= Toscani in Italia; YRI= Yoruba in Ibadan, Nigeria.

Table 3: Multiple regression with random SNPs and GWAS hits.

Replication of Scores by Piffer (2015)

Piffer [2] calculated polygenic and factor scores for 1000

Genomes population using the data available at the time. The correlation between the present study's EA and intelligence factors and Piffer's 2015 polygenic score are very high (r=0.96-0.99).

ALFRED

The 9 quasi-replicated EA SNPs and the 18 intelligence hits were searched in ALFRED (52 populations). Four of them were found. Rs9320913 (a hit in [10-12]) was not found but it was in LD (r=0.95) with Rs1906252 [2]. A factor analysis was carried out. Factor loadings are reported in (Table 4).

SNP	Loading
rs10236197_T	-0.218
rs1906252_T	0.533
rs13010010_T	0.868
rs11588857_A	0.790

Table 4: Factor loadings.

Polygenic score and factor scores are reported in (Table 5).

	Population	Factor	PS
Africa	Bantu SA	-1.454	0.063
Africa	Bantu Kenya	-1.381	0.168
Africa	San	-1.488	0.188
Africa	Biaka	-1.369	0.198
Africa	Mbuti	-1.415	0.204
Africa	Yoruba	-1.270	0.168
Africa	Mandenka	-1.153	0.150
N. Africa	Mozabite	-0.768	0.294
Asia	Bedouin	-0.156	0.378
Asia	Druze	0.254	0.416
Asia	Palestinian	-0.071	0.405
Europe	Adygei	0.257	0.360
Europe	Basque	-0.088	0.388
Europe	French	0.217	0.408
Europe	Italians_C	0.404	0.405
Europe	Italians_N	0.437	0.413
Europe	Orcadian	0.753	0.493
Europe	Russians	0.073	0.393
Europe	Sardinian	-0.225	0.315
Asia	Burusho	0.151	0.390
Asia	Kalash	0.475	0.375
Asia	Pashtun	-0.426	0.345
Asia	Mongolian	1.358	0.488
Asia	Balochi	0.055	0.363
Asia	Brahui	-0.334	0.370
Asia	Hazara	0.506	0.458
Asia	Sindhi	-0.438	0.325
EastAsia	Dai	0.987	0.463
EastAsia	Daur	1.246	0.540
EastAsia	Han	0.936	0.410
EastAsia	Hezhe	0.980	0.488
EastAsia	Japanese	1.018	0.410
EastAsia	Koreans	1.127	0.471
EastAsia	Lahu	0.877	0.388
EastAsia	Miao	1.078	0.525
EastAsia	Naxi	0.113	0.363
Asia	Oroqen	0.445	0.325
EastAsia	She	0.737	0.488
EastAsia	Tu	0.828	0.450
EastAsia	Tujia	1.507	0.463
EastAsia	Uyghur	0.566	0.475
EastAsia	Xibe	0.802	0.375
EastAsia	Yi	1.190	0.413
EastAsia	Cambodians, Khmer	0.340	0.346
Oceania	Papuan New Guinean	-0.569	0.348
Oceania	Melanesian, Nasioi	-0.533	0.334
Siberia	Yakut	0.311	0.432
NorthAmerica	Pima, Mexico	-1.312	0.320
NorthAmerica	Maya, Yucatan	-1.300	0.204
SouthAmerica	Amerindians	-1.366	0.260
SouthAmerica	Karitiana	-1.530	0.200
SouthAmerica	Surui	-1.382	0.245

Table 5: Factor and polygenic scores for HGDP (ALFRED) populations.

Factor scores broken down by sub-continent are reported in (Table 6).

Sub-continent	Factor	PGS
Africa	-1.287	0.179

M East	0.009	0.400
Europe*	0.293	0.408
W Asia	-0.002	0.375
E Asia**	0.959	0.450
Oceania	-0.551	0.341
SE Asia	0.340	0.346
America	-1.378	0.246
Siberia	0.311	0.432
North Africa	-0.768	0.294
*Sardinia is not included in the European group due to its status as genetic outlier and higher similarity to Middle Eastern populations. ** Includes Mongolia.		

Table 6: Factor scores by continent.

The correlation between factor scores distances and Fst distances was r=0.462 (N=1176). The correlation between geographic distance from Addis Ababa and factor scores was r= - 0.483 (N=49)

Discussion

Two methods were used to estimate polygenic selection and related genotypic differences in intelligence and educational attainment between populations.

The calculation of population-level polygenic scores (average allele frequencies with positive GWAS beta) is a promising and quick approach. However, it is a-theoretical relative to evolutionary processes. Factor analysis of allele frequencies is another method, whose goal is to detect a hypothetical signal of polygenic selection which accounts for covariance among frequencies of several alleles associated with the same trait. A drawback of this method is it can be performed only on small sets of SNPs, because using large sets would excessively reduce the observations to variable ratio, hence making the results of factor analysis less stable.

The polygenic score obtained from 9 quasi-replicated SNPs is a good candidate for estimating selection strength on educational attainment: it outperforms (in predicting average population phenotypic intelligence) 99% of the polygenic scores obtained from random SNPs (Monte Carlo p= 0.011): that is, over a total of 819 runs, a correlation coefficient equal to or higher than 0.88 occurred 8 times, producing p= 0.011. It has high predictive power (Beta= 0.88), being robust to tests controlling for spatial autocorrelation (Table 2). The larger polygenic score (161 SNPs) also had a high predictive power (Beta= 0.8, after controlling for spatial autocorrelation), outperforming 98% of the random polygenic scores. A recent GWAS produced 18 independent genomic hits for intelligence.

The results produced by the intelligence GWAS were positive only with regards to the factor scores. Monte Carlo simulations showed that the intelligence polygenic scores failed to predict population IQ significantly better than random SNPs.

However, factor analysis yielded factor scores that strongly predicted population IQ (r=0.9). The significance of this correlation was ascertained via Monte Carlo simulation, and found to be high (Monte Carlo p= 0.004). Moreover, the factor scores were also highly correlated to the 9 SNPs educational attainment PS (r= 0.89) and to the 161 SNPs PS (r= 0.88). Monte Carlo simulations show that this inter-correlation was significant (p=0.016 and 0.021, respectively).

A factor analysis was carried out on the 9 educational attainment SNPs and the factor scores were found to be almost identical (r=0.96) to the intelligence factor scores. This correlation had MC p= 0.023. Hence, support was found for the hypothesis that educational attainment and intelligence are genetically correlated at the group level.

The EA polygenic and factor scores and the intelligence factor were robust to tests of spatial autocorrelation (table 3). When IQ was regressed on the factor scores and Fst distances, the latter lost all the predictive power, whereas the former had high Beta (0.81-0.84). More strength to the present findings is given by the high replicability of the same polygenic model from a previous publication [2]. Piffer [2] calculated polygenic and factor scores for 1000 Genomes population using the data available at the time. The correlation between the present study's EA and intelligence factors and Piffer's 2015 polygenic and factors score are very high (all of the 4 correlations range between r=0.96-0.99). This is remarkable, since the present study included 2 new educational attainment GWAS [10,11] and a new intelligence GWAS [12], comprising much larger samples.

The results were replicated using a larger sample of populations (N=52), which showed similar population and continental rankings of factor scores (Tables 5,6). A positive correlation with Fst distances was found, but a negative one with distance from Eastern Africa. The latter finding casts doubt on the hypothesis that evolutionary novelty (operationalized as geographic distance from the ancestral African environment) was a force behind the evolution of general intelligence [21]. However, this dataset had much lower coverage and only a handful of SNPs could be retrieved.

The much higher East Asian scores (Table 6 and 7) also suggests that strong selection pressure on East Asians continued after the East Asian-Native American split, about 15 kya at the earliest (the earliest estimate of a migration across the Bering strait into the Americas). It is possible that the extremely low population density in the Americas reduced intraspecific competition, hence selection pressure on cognitive ability was lower.

A limitation of the present study is its reliance on estimates of population IQ as the phenotypic variable, which are not

perfectly accurate, besides reflecting environmental and economic differences between populations.

The low amount of variance explained at the individual level is not a fatal issue. Indeed, predicting group-level variance is different from predicting variance within a group. The present approach maximizes signal by focusing on the average signal produced by each SNPs, rather than the sum of the signal produced by all the SNPs. Using more SNPs would include more noise in the data, as SNPs with lower significance likely carry out too much noise. One could argue that weighting each SNP by its regression coefficient or p value could allow us to assign less weight to the noisier SNPs, hence getting rid of this issue. A similar approach was used in [20]. However, this approach assumes a linear relationship between population level noise and GWAS significance, which has never been demonstrated. In fact, it is possible that the relationship between GWAS effect/p value and population-level signal is exponential and concentrated among a handful of SNPs.

The effect of LD decay on comparison of risk alleles between populations is still unclear. Since most GWAS hits are actually tag SNPs, decay in LD implies that the causal SNPs will be less efficiently flagged by the tag SNPs until the tag SNPs will resemble a sample of random SNPs. It is sensitive to coverage and in older studies using low coverage genomic data (e.g. 1000 Genomes phase 1), it was found to reduce the reproducibility of findings [21]. However, contemporary GWAS use higher coverage data (e.g. 1000 Genomes phase 3), hence this issue is less important.

Moreover, simulations found that the effect of LD decay on true causal variants was null to negligible [22]. The present study, by focusing only on the most significant hits, increased the likelihood of hitting on or very close to causal variants, hence reducing the artifact of LD decay. Furthermore, LD decay is expected to create noise and not to produce a bias necessarily in a direction that favours the hypothesis of this study [23].

Since the frequency of the average SNP allele is 50%, the tag SNPs will tend to converge towards an average frequency of 50%, with increasing LD decay. The implication of this for our analysis is that our estimates of non-European polygenic scores will be inflated, because the average frequency of the alleles with positive effect is lower than 50% among Europeans (e.g. 44% and 45% for the 9 EA and 18 intelligence hits), and LD decay pushes the polygenic scores up towards the background frequency of 0.5 in the other groups. Methods to tackle the issue of LD decay are currently being developed [24]. It is not clear how LD decay affects factor scores, but their high reproducibility and predictive validity yield credibility to the results.

Conclusion

This study successfully replicated the findings from previous ones using more powerful and reliable statistical tests using data obtained from large, recent GWA studies. However, some issues are still unresolved (e.g. LD decay) and it is to be hoped that future GWAS will be carried out on non-European populations in order to have a more cross-racially valid phenotypic prediction. Such a feature is not impossible and was recently achieved on skin pigmentation: a set of 36 SNPs was identified that was highly predictive of skin color within and across different super-populations [25].

Although the genetic architecture of skin color is different from that of cognitive abilities, this should serve as an encouragement to carry out studies on racially diverse populations with the aim to identify universally-valid markers of intelligence.

References

1. Piffer D (2013) Factor analysis of population allele frequencies as a simple, novel method of detecting signals of recent polygenic selection: The example of educational attainment and IQ. Mankind Quarterly 54: 168-200.

2. Piffer D (2015) A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. Intelligence 53: 43-50.

3. Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Current biology 20: 208-221.

4. Berg JJ and Coop G (2014) A population genetic signal of polygenic adaptation. PLoS Genetics 10: e1004412.

5. Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, et al. (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nature Genetics 44: 1015-1101.

6. Zoledwieska M, Sidore C, Cucca F, Chiang CWK, Sanna S, et al. (2015) Height-reducing variants and selection for short stature in Sardinia. Nature Genetics 47: 1352-1356.

7. Robinson RR, Hemani G, Medina-Gomez C, Mezzavilla M, Esko T, et al. (2015) Population genetic differentiation of height and body mass index across Europe. Nature Genetics 47: 1357-1362.

8. Piffer D (2016) Evidence for Recent Polygenic Selection on Educational Attainment Inferred from GWAS Hits. Preprints 2016, 2016110047.

9. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, et al. (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 340: 1467-1471.

10. Davies G, Marioni RE, Liewald DC, Deary IJ, Craddock N, et al. (2016) Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Molecular Psychiatry 21: 758-767.

11. Okbay A, Beauchamp JP, Fontana MA, Lee J, Pers TH, et al. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533: 539-542.

12. Sniekers S, Stringer S, Watanabe K, Posthuma D, Coleman JRI, et al. (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence, Nature Genetics 49: 1107-1112.

13. Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, et al. (2014) The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. PNAS 111: 15273-15278.

14. Machiela MJ and Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31: 3555-3557.

15. Davison AC and Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge, United Kingdom 1997.

16. Pers TH, Timshel P, Hirschhorn JN (2014) SNPsnap: a Web-based tool for identification and annotation of matched SNPs. Bioinformatics 31: 418-420.

17. Handley LJ, Manica A, Goudet J, Balloux F (2007) Going the distance: human population genetics in a clinal world. Trends in Genetics 23: 432-439.

18. R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

19. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research. 27: 209-220.

20. Kanazawa S (2008) Temperature and evolutionary novelty as forces behind the evolution of general intelligence. Intelligence 36: 99-108.

21. Domingue BW, Belsky DW, Conley D, Mullan Harris K, Boardman JD (2015) Polygenic Influence on Educational Attainment. AERA Open 1: 1-13.

22. Zanetti D, Weale ME (2016) True causal effect size heterogeneity is not required to explain trans-ethnic differences in GWAS signals. bioRxiv.

23. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, et al. (2017) Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. The American Journal of Human Genetics 100: 635-649.

24. Piffer D (2017) LD decay.

25. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, et al. (2017) Global skin color prediction from DNA. Hum Genet 136: 847-863.