Divergent selection on height and cognitive ability: evidence from both genetic distance (Fst) and polygenic scores

Davide Piffer*

OpenPsych

Abstract

Tests of selection based on population differentiation were performed on two highly polygenic traits important for success and quality of life: body height and educational attainment (EA). Polygenic scores (PGS) of EA and height, computed across three public genomic databases revealed differences between populations (1000 Genomes, HGDP, gnomAD) that matched the average IQ and height of ethnic groups (r ∼0.9). A moderately strong correlation between latitude and EA PGS (r= 0.67) implies the effect of climate (seasonality or winter temperature) on selection for cognitive ability. The effect of latitude was reduced (β= 0.28) but remained significant after adding the sub-continental group variable to the regression model. The global Fst index revealed population differentiation at height and EA loci, significantly deviating from random SNPs. This is indicative of directional selection pressures with different strengths across groups. Substantial Linkage Equilibrium (LD) Decay between Africans and Europeans was found (r= 0.6) but there was no correlation between LD decay and population differences in polygenic scores for EA (r= 0.015, p= 0.45), and slight inflation of height PGS difference due to LD decay (r= -0.04, p= 0.0315). Selecting the SNPs most robust to LD decay (r>0.8) resulted in larger PGS gaps for EA, but smaller for height. Finally, it is shown that PGS differences are more sensitive to the significance of GWAS loci than Fst, reflecting the major limitations of Fst as an index of selection.

Keywords: Educational attainment; polygenic score; GWAS; polygenic adaptation; height

1 Introduction

Many studies over the last decade have established the highly polygenic nature of many human traits (Loh et al., 2015; Shi et al., 2016; Boyle et al., 2017). This discovery was aided by technical advances in the science and technology of genomics, which enabled genome-wide association studies (GWAS) on large samples to detect thousands of variants with tiny effect on human traits. This genetic architecture contrasts with the classical Mendelian model where effects are driven by rare variants with large effect.

Despite these advances, polygenic adaptation has proven difficult to disentangle because causal variants only undergo subtle changes in allele frequency. Moreover, the arrays used by GWAS usually tag causal variants, meaning that they only identify variants that are in linkage disequilibrium with the causal variant. Whilst the effect of causal variants is generally homogeneous across ethnic groups (Ishigaki et al., 2020),

tag variants have decreased predictive power in populations that are ancestrally different from the GWAS reference sample, due to a phenomenon known as Linkage Disequilibrium decay. LD decay is due to recombination events that shuffle the genetic material (Vos et al., 2017). Another issue that affects GWAS and selection methods relying on GWAS summary statistics is population stratification, which can inflate signals of selection due to co-variance between genes and the environment.

Methods to detect directional selection usually exploit population differentiation in allele frequencies (Piffer, 2013; Berg & Coop, 2014). Other methods are based on within-population variation: selection scans based on singleton density score (SDS) (Field et al., 2016) or tests that analyze the correlation between derived allele frequencies (DAF) and GWAS effect size and direction (Uricchio et al., 2019).

Other tests rely on the correlation between allele frequencies and environmental variables, such as annual temperature or rainfall, or proxies for climate such as latitude (Limborska et al., 2002; Eisenberg et al., 2010).

* Department of Biology, Tuebingen University, 72074 Tuebingen, Germany. Corresponding author, E-mail: pifferdavide@gmail.com

This paper will use tests relying on between-population variation and correlation with latitude to identify signals of polygenic adaptation on two highly polygenic traits: Educational attainment and height. The correlation of polygenic scores with average population trait (phenotype) and Qst-Fst (Whitlock & Guillaume, 2009) test are complementary approaches to detect polygenic selection. The correlation between polygenic scores and environmental variables such as latitude has a long tradition and it has become part of standard test of selection packages (Berg & Coop, 2014).

Population differentiation at education-related genetic variants has been shown by previous studies (Piffer, 2013, 2015, 2019). Recently a study found strong directional selection on educational attainment in Britain over the last 2k years (Stern et al., 2021). This trend has reversed in many countries after the industrial revolution, and molecular evidence for this phenomenon was provided by Kong et al. (2017) which showed that in Iceland, the EA PGS has been decreasing over time because people with higher EA polygenic scores tend to have fewer children.

Education is highly genetically correlated to intelligence (r>0.69), and can be used as a proxy for cognitive abilities (Davies et al., 2018; Hill et al., 2019; Savage et al., 2018). However, other non-cognitive skills contribute to educational attainment, hence the selection signal might reflect selective pressure on other traits. Recently, a study found genetic associations with EA independent of cognitive abilities. Some traits such as extraversion, agreeableness and conscientiousness were positively correlated to the non-cognitive factor but negatively to intelligence (Demange et al., 2021).

Unfortunately the resulting non-cognitive factor retained some variance in common with intelligence (although to a much smaller extent than the cognitive PGS), preventing us from treating the two factors separately in the population-level analysis.

The existence of non-cognitive traits whose genetic correlation with intelligence and education go in the opposite direction makes it problematic to use EA as a proxy for IQ. For this reason, this study will use the summary statistics from the multi-trait analysis (MTAG) of education with cognitive performance, highest math class completed and self-reported math ability (Lee et al., 2018), which likely reflect the contribution of cognitive genetics more than the pure EA phenotype ("Years of Education"). Indeed, the MTAG PGS explained 11-13 % of the variance in EA and 7-10 % in cognitive performance (Lee et al., 2018).

These two traits were chosen because they are:

1) socially relevant since they both affect success in life (academic, occupational and mating success);

2) highly polygenic and

3) GWAS are available for very large samples.

Finally, two novel measures of selection are introduced:

1) the correlation of GWAS allele frequencies across pairs of populations and

2) the ratio between allele frequency difference and mean absolute allele frequency difference.

2   Methods

Polygenic scores were computed using GWAS SNPs meeting the standard significance threshold ($p< 5 \times 10^{-8}$) and weighed by effect size. The inclusion criteria for GWAS were sample size and predictive power: the GWAS with the largest sample size and predictive power were used to compute polygenic scores. For education, the Lee et al. (2018) GWAS met these criteria: it had the highest predictive power (~9 %) and largest sample size (1.1 million) among GWAS of education and cognition. The EA MTAG (henceforth, "EA3") polygenic score was chosen because it encompassed several cognitive abilities and had the largest predictive power; for height, the GWAS meta-analysis comprising 700,000 individuals by Yengo et al. (2018) had the largest sample size, but the highest predictive power (42.8 %) was achieved by Chung et al. (2019), albeit with a smaller sample size (N= 456,837). The latter was used for computing polygenic scores, and the former to compute Fst distances and LD decay, because (unlike Chung et al. (2019)) the publicly available statistics provided the necessary p-values to perform LD clumping and p-value thresholding.

Polygenic scores were computed using the three largest publicly available population genetics datasets: 1000 Genomes ("1KG"), gnomAD (https://gnomad.broadinstitute.org/) and the Human Genome Diversity Project (HGDP) (ftp://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/).

Average height for populations was obtained from Wikipedia (https://en.wikipedia.org/wiki/Average_human_height_by_country) and was available for 12 populations in 1000 Genomes. Only male average height was used, because statistics for females are absent for many countries. For countries with more than one value, the most recent statistics was used. The HGDP sample included mostly tribes that were not represented in national height statistics for which it would be difficult to get a precise estimate,

hence they were reported without attempting to compute the correlation with phenotypic height. The average height for gnomAD was obtained from National Health Statistics (Fryar et al., 2018) for ethnic groups in the US represented by the gnomAD samples (Hispanic, East Asian, European (non Finnish), African-American). The average height for Finns was obtained from Wikipedia. The average height of Ashkenazi Jews was obtained from a recent study on the genetics of height among a sample consisting mostly of Ashkenazi Jews (Zeevi et al., 2019).

Latitude values for HGDP populations were obtained from Rosenberg et al. (2005) (see: https://rosenberglab.stanford.edu/data/rosenbergEtAl2005/rosenbergEtAl2005.coordinates.txt). The 1000 Genomes and gnomAD contained samples that did not overlap with HGDP, but these were mostly collected from non-indigenous groups. Thus, only latitude for the 53 HGDP samples was used.

The other two datasets (1KG and gnomAD) contained many samples (Ashkenazi, Amish, Utah Whites, African Americans, etc.) which were collected from non-indigenous groups, hence the correlation with latitude was computed only for the HGDP dataset.

The random SNPs were matched by MAF (minor allele frequency) and LD (with a threshold r^2=0.1) using SNPsnap (Pers et al., 2015).

The absolute allele frequency difference was calculated as the mean absolute difference of the allele frequencies Fst and allele frequencies were calculated using Plink 1.9 and polygenic scores for each population were computed using R (version 3.6). Code and data are available on OSF: https://osf.io/6dvfc/

3 Results

3.1 Height PGS

The polygenic scores for height were calculated using Chung et al. (2019) GWAS, using the effect size in the Lasso+CTPR meta-analysis, which provided the best predictive power. These PGS showed substantial inter-population variation. The highest values were obtained by northern Europeans (Finns, Orcadians, Northern Europeans from Utah) and the lowest by Native Americans (Karitiana, Surui) and Southeast Asians (Dai, Vietnam).

There was a weak, positive correlation (r= 0.321) between height PGS and latitude (Figure 1).

The height PGS in the 1000 Genomes (Figure 2) showed higher scores for Europeans and Africans, and lower for Latin Americans and East Asians.

The PGS had a strong correlation with average population height (r= 0.923) (Figure 3).

In the gnomAD dataset, the correlation between height and PGS was 0.949 (Figure 4).

3.2 Education

Among the HGDP populations, there was a positive correlation (r= 0.675) between EA3 PGS and latitude (Figure 5).

3.2.1 HGDP

East Asian ethnic groups had positive residuals, suggesting the presence of superpopulation-specific factors. Hence, a multiple linear regression was run, adding sub-continent as a categorical variable. The standardized coefficients are shown in Table 1 with Africans (omitted) as the reference level.

Table 1: Multiple regression (PGS ~ Latitude + superpopulation): standardized coefficients.

Variable	Beta	S.E.	p
Latitude	0.28	0.11	0.014
Amerindian	0.42	0.28	0.145
C. Asian	1.32	0.28	$2.25*10^{-5}$
East Asian	1.96	0.26	$1.79*10^{-9}$
European	1.28	0.35	0.0007
Oceanian	0.96	0.40	0.019
Middle Eastern	1.19	0.36	0.002

3.2.2 Monte Carlo simulation

1000 random polygenic scores were generated to compute the null distribution of correlation coefficients. The correlation between the clumped EA3 PGS and latitude (significantly deviated from the null (Z= 3.29, p= 0.0009) (Figure 6).

3.2.3 1000 Genomes

The EA PGS in the 1000 Genomes (Figure 7) showed higher scores for East Asians and Europeans, and lower for Latin Americans and Africans.

Among the 1000 Genomes populations, the correlation between EA3 PGS and average IQ was r= 0.903 (Figure 8).

3.2.4 gnomAD

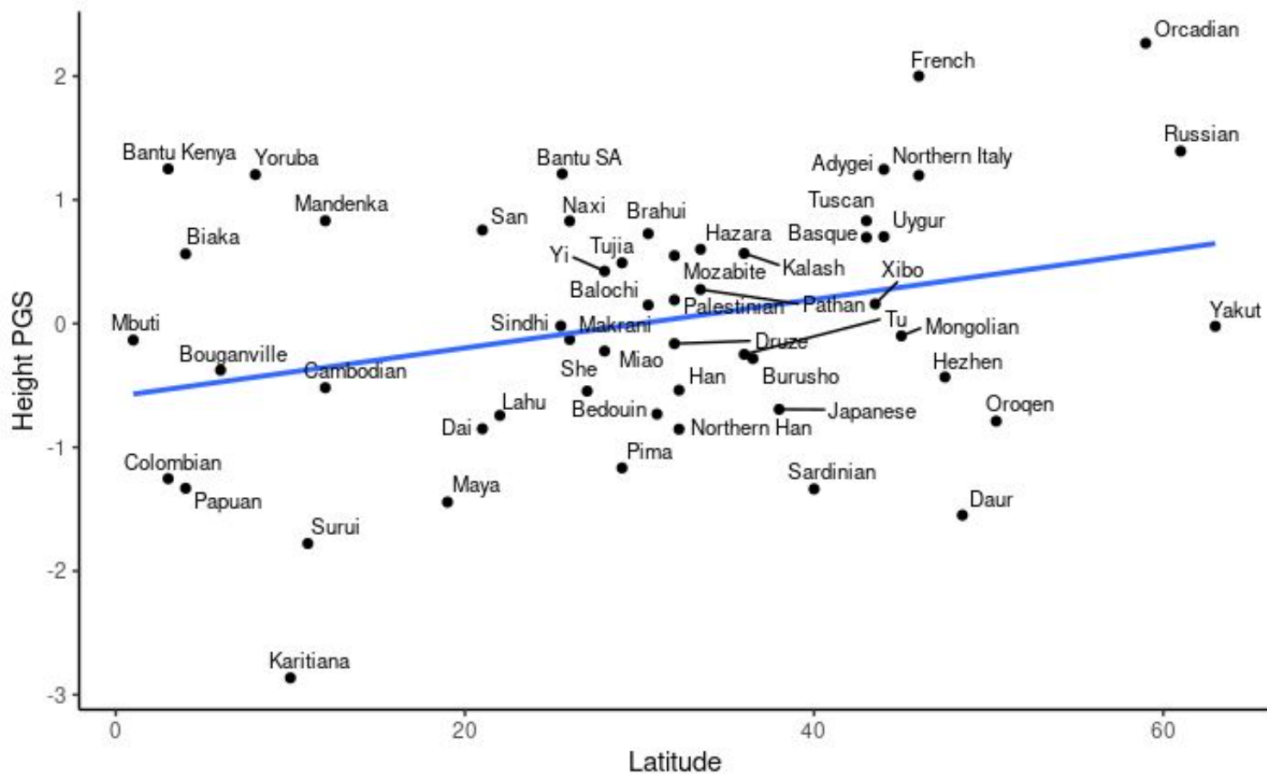The correlation between IQ and EA3 PGS was r= 0.985 (Figure 9).

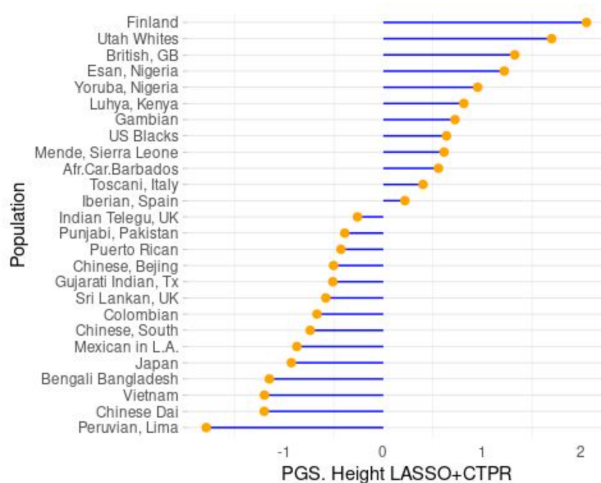**Figure 1:** Scatterplot (Height PGS x Latitude): HGDP populations.

Figure 2: Height PGS: 1000 Genomes.

3.2.5   Correlation between height and education (PGS and average height)

The population-level correlation between EA3 and height PGS was close to 0 (r= 0.02, -0.304, 0.03 in gnomAD, 1000 Genomes and HGDP, respectively).

3.3   Global Fst

A version of the Qst-Fst test, known as Fst enrichment test (Guo et al., 2018) was used to detect divergent selection. A more powerful version is used here, which

avoids the issue of multiple comparisons found in pairwise tests: global Qst-Fst. The trait-associated SNPs were matched (after LD clumping) to random SNPs stratified by MAF and not in close LD (r< 0.1). This process was repeated 1000 times to generate a distribution of mean Fst under neutrality.

The GWAS by Yengo et al. (2018) was used to compute Fst values for height-associated SNPs because the Chung et al. (2019) GWAS summary statistics lacked the p value needed to perform LD clumping.

In 1KG, the height PGS had higher average Fst than random SNPs, significantly deviating from Fst values of random PGS (Figure 10).

A similar deviation from random SNPs was observed in the HGDP dataset (Figure 11).

The global Fst enrichment test yielded positive results for education too (Figure 12 and Figure 13).

The Fst values for random and GWAS SNPs and the Z scores + empirical p values are reported in Table 2.

3.4   Controlling for LD decay

The SNPs that reached genome-wide significance (P< 5×10^{-8}) in the largest GWAS of educational attainment (Lee et al., 2018) were selected. Those with minor allele frequency (MAF) <0.01 in 1KG database among Europeans (CEU) and Africans (YRI) were
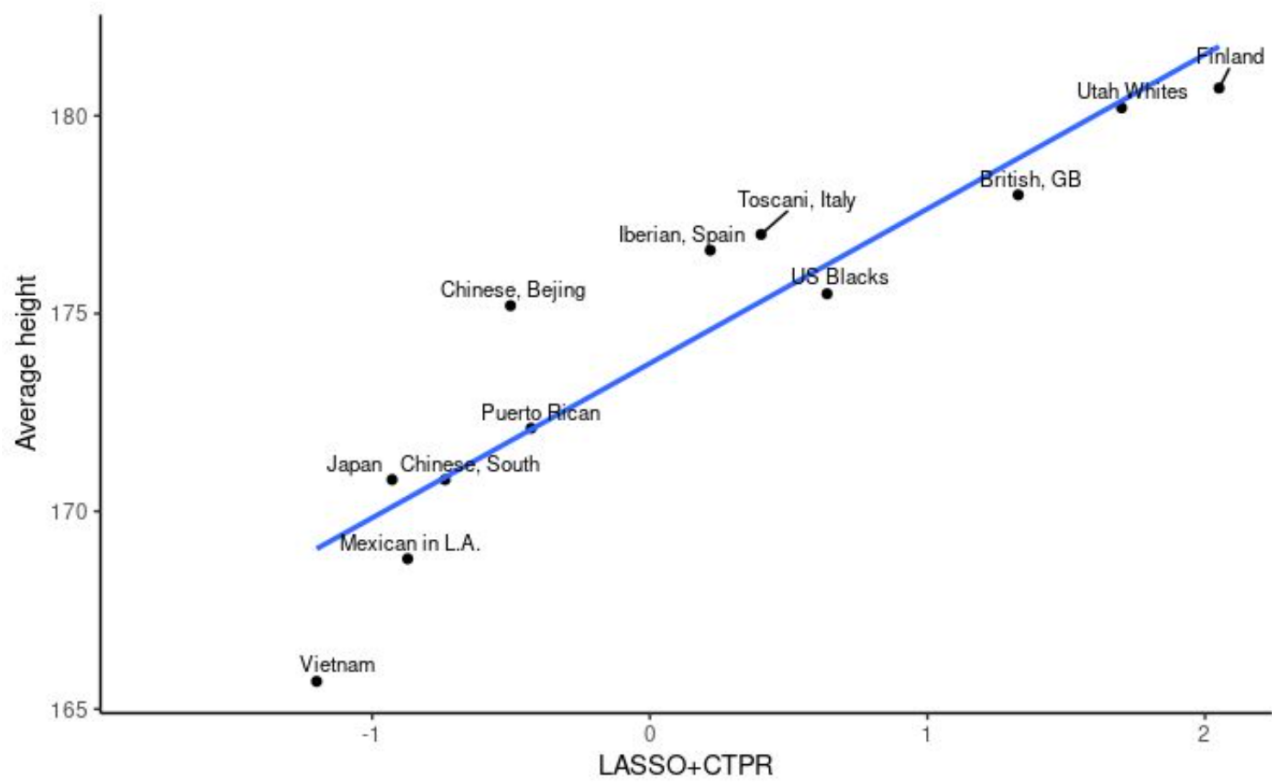
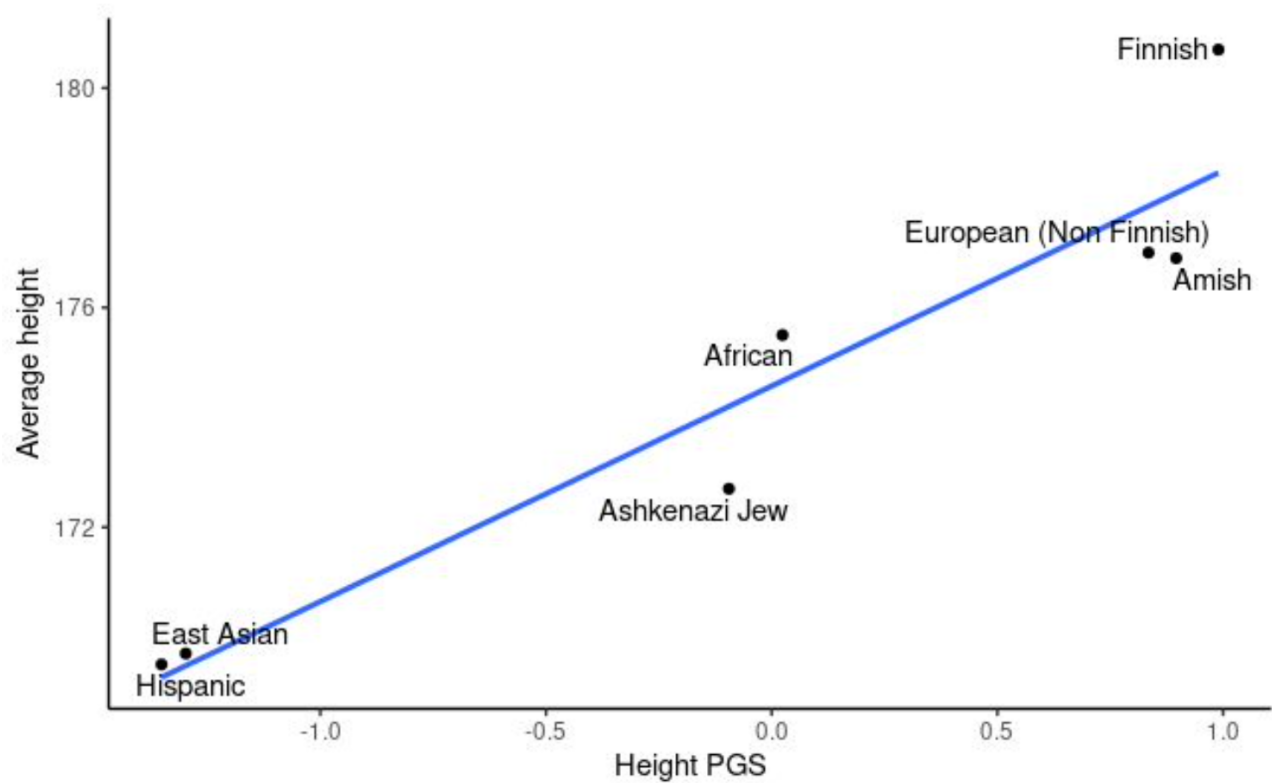**Figure 3:** Scatterplot (Height PGS x average height).
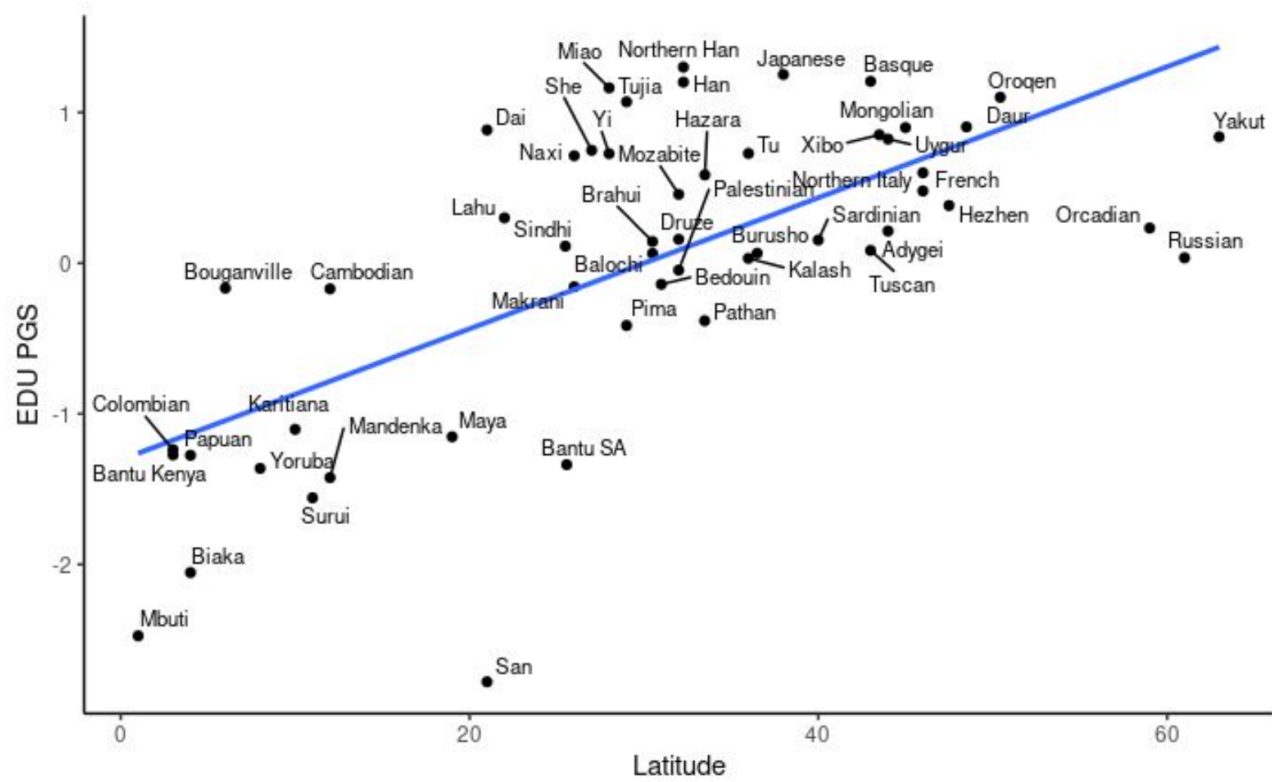

Figure 4: Scatterplot (Height x PGS): gnomAD.

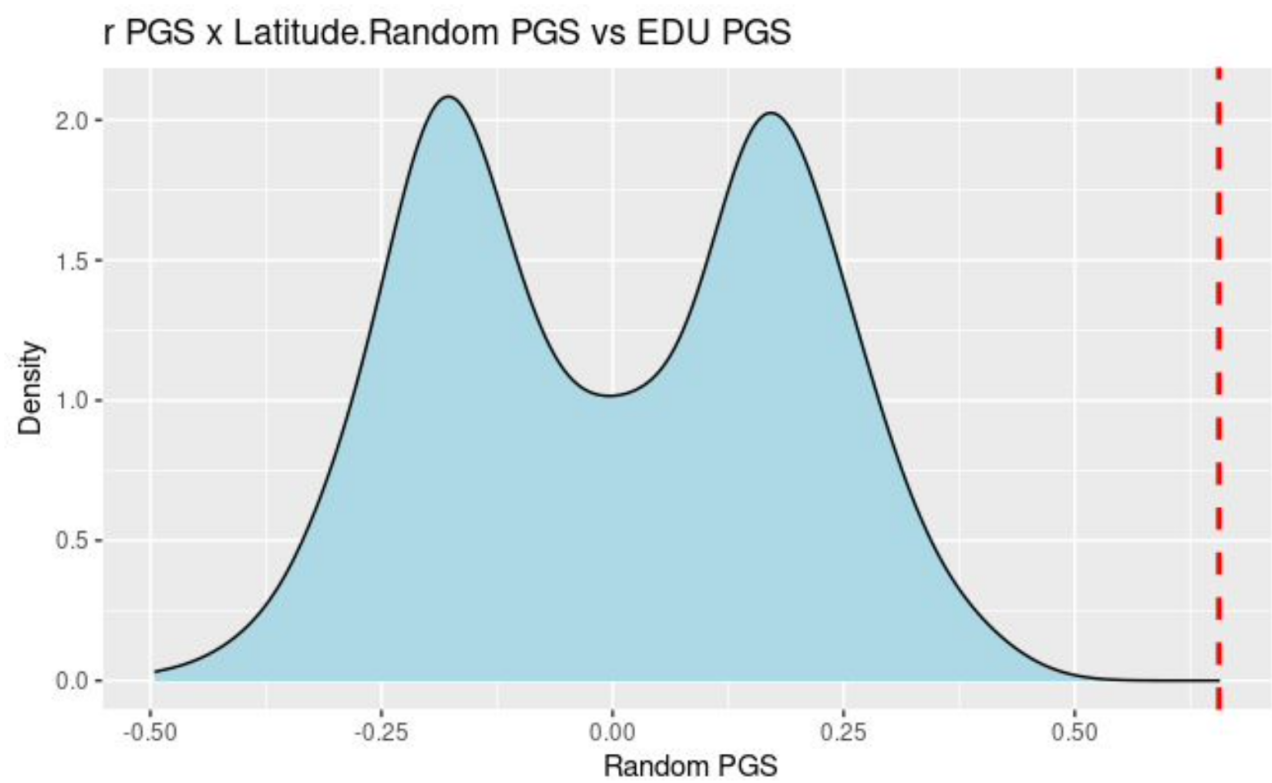**Figure 5:** Scatterplot (EA3 PGS x Latitude): HGDP populations.
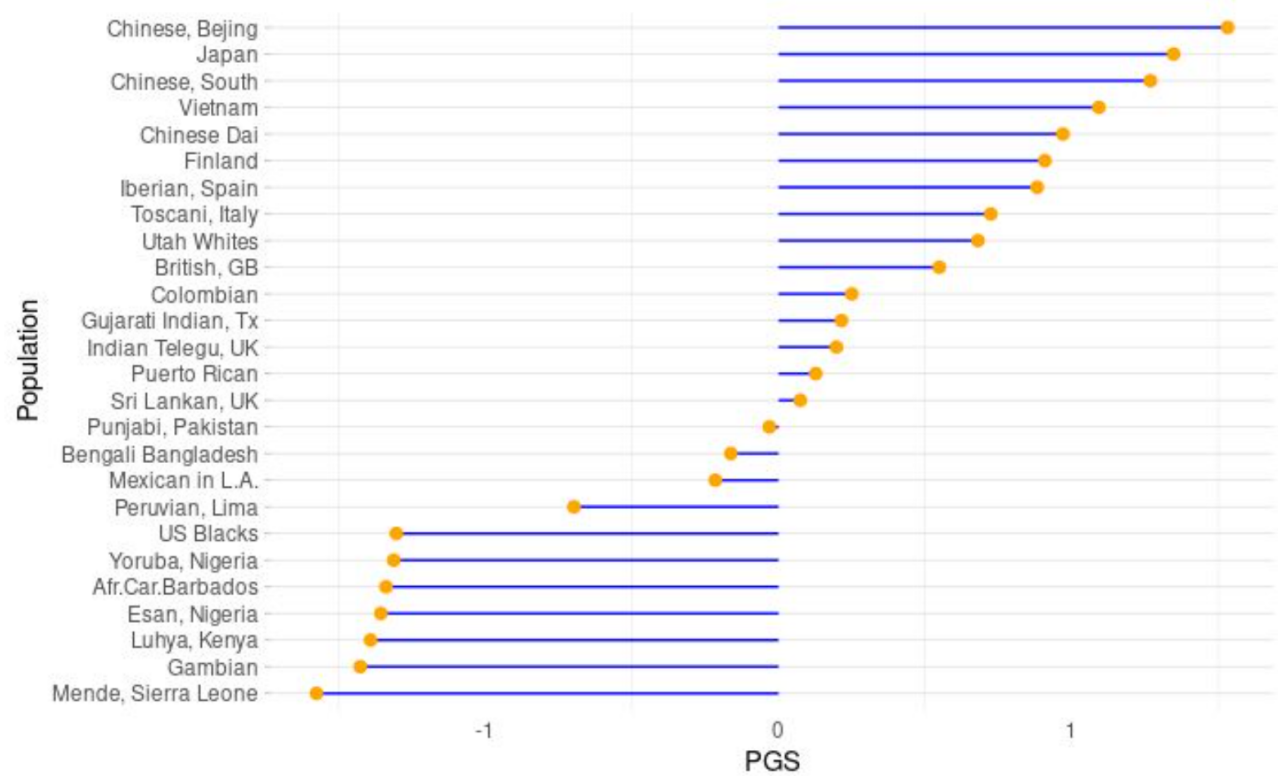

Figure 6: Density plot (correlation with latitude).
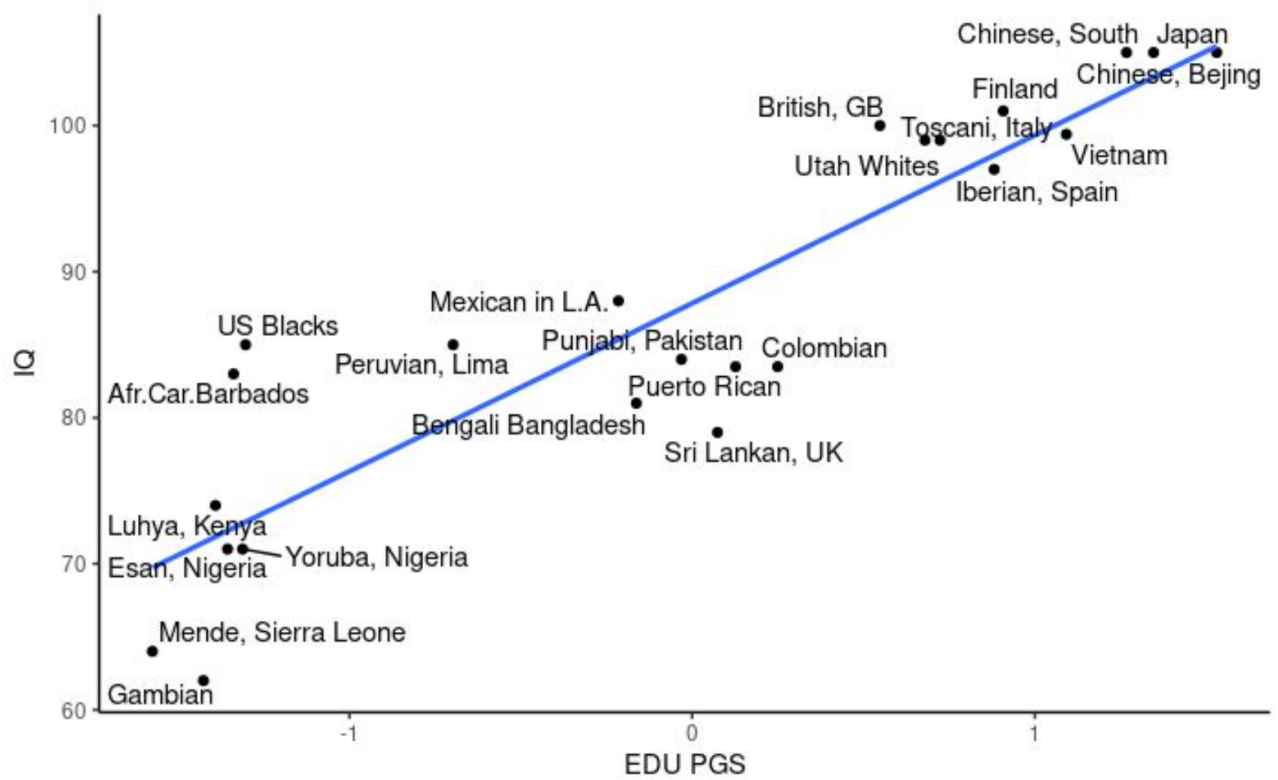
Figure 7: EA3 PGS: 1000 Genomes.

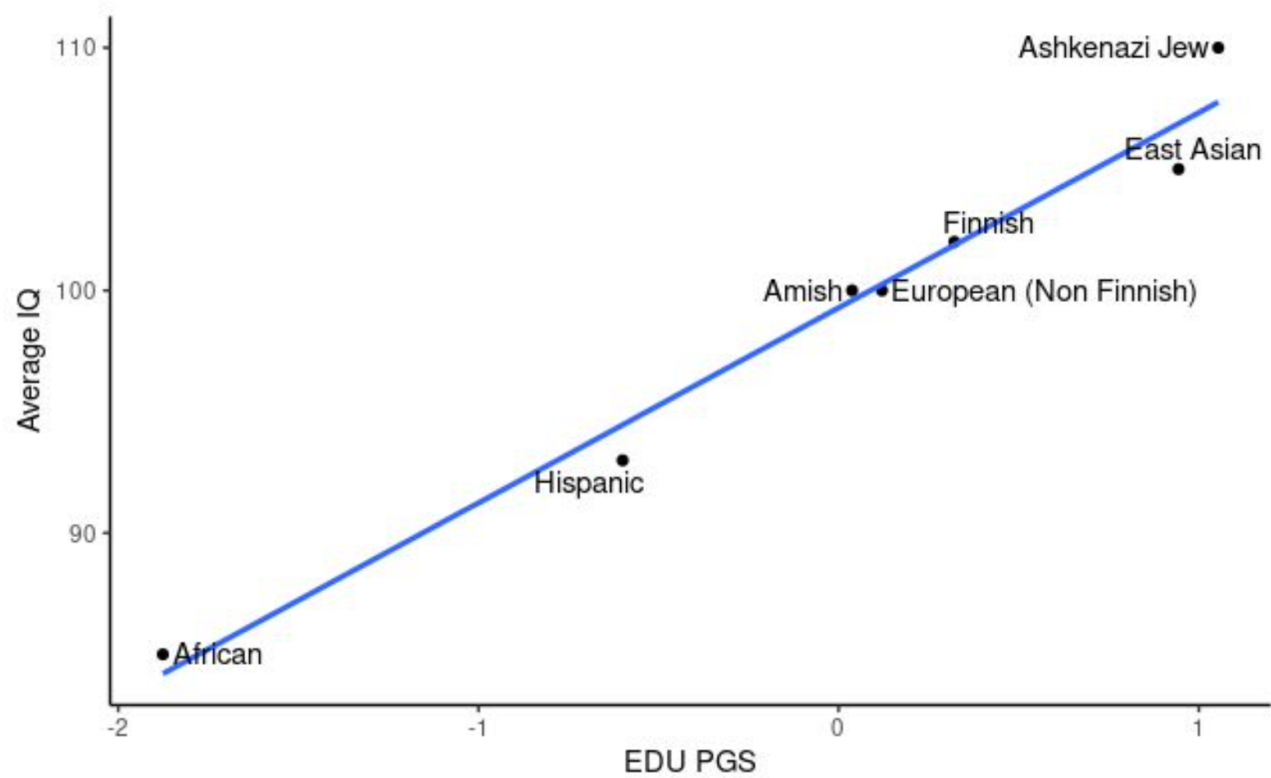**Figure 8:** Scatterplot (EA3 PGS x IQ): 1000 Genomes.

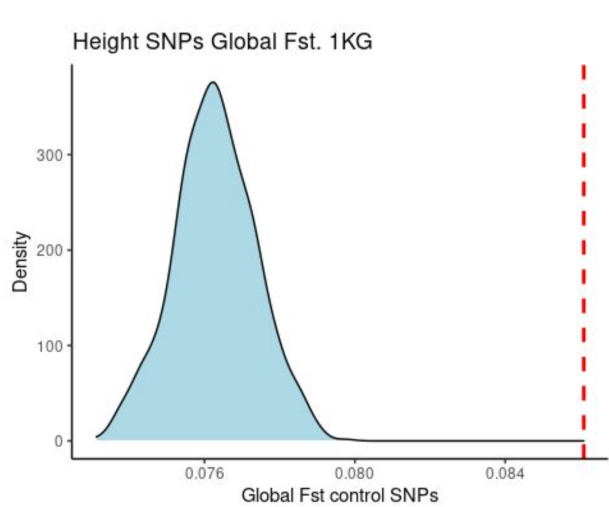**Figure 9:** Scatterplot (EA3 PGS x IQ): gnomAD.

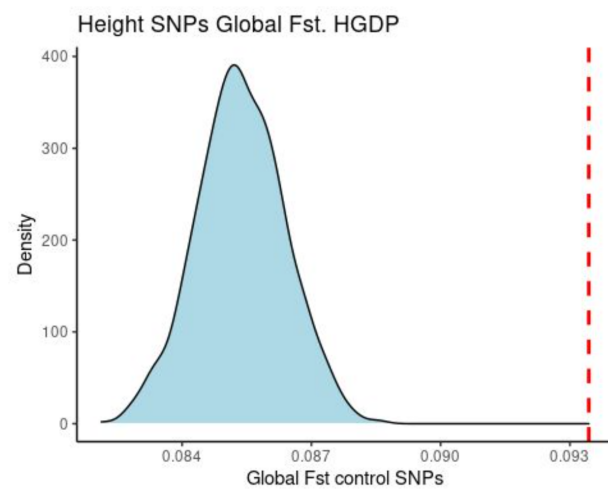**Figure 10:** Global Fst for height SNPs and random SNPs. 1000 Genomes.

Figure 11: Global Fst for height SNPs and random SNPs. HGDP.

Table 2: Fst, Z score and p values for height Fst enrichment test.

	Fst GWAS	Fst random	Z	p
Height 1KG	0.086	0.076	9.04	0.001
Height HGDP	0.089	0.084	3.679	0.0009
EA3 1KG	0.080	0.076	3.11	0.002
EA3 HGDP	0.093	0.085	7.945	0.0009

removed. There were 2,596 SNPs satisfying these criteria. To explore the different LD patterns across populations, the SNPs were uploaded to LDlink, and all variants -/+ 500 Kb of the query variant with a pairwise R2 value greater than 0.01 were downloaded using CEU and YRI as reference populations. There were 4,696,863 and 5,739,447 variants for CEU and YRI, respectively. The two files were merged and 1,680,781 overlapping SNPs were retained.

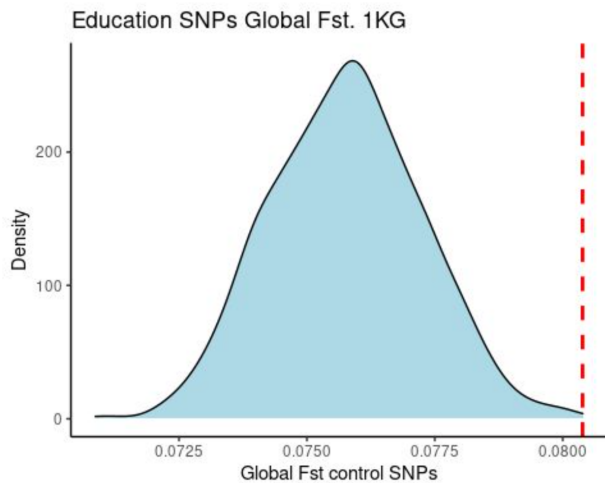For each query variant, the correlation between the

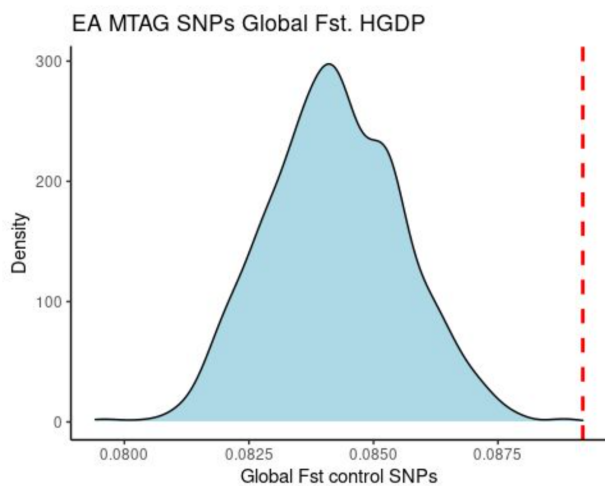**Figure 12:** Global Fst for EA3 SNPs and random SNPs. 1000 Genomes.

Figure 13: Global Fst for EA3 SNPs and random SNPs. HGDP.

pairwise R2 values for CEU and YRI was computed. The correlation coefficient was used as an index of differential LD decay across CEU and YRI relative to the query variant. That is, the higher the correlation between the CEU and YRI R2 values, the lower the amount of trans-ethnic LD decay.

For each SNP, the PGS for CEU and YRI was computed, as well as the pairwise Fst.

The average correlation coefficient between the CEU and YRI R2 values across the 2596 SNPs was r=0.608. This shows the presence of a moderate amount of LD decay.

There was a statistically significant PGS difference between CEU and YRI. Welch Two Sample t-test: means: 50.312 % - 47.995 % respectively, t= 3.134, 95 % C.I 0.008 - 0.037, p= 0.001.

There was no correlation between LD decay and the CEU-YRI PGS difference (r= 0.015, p= 0.451) (Fig-
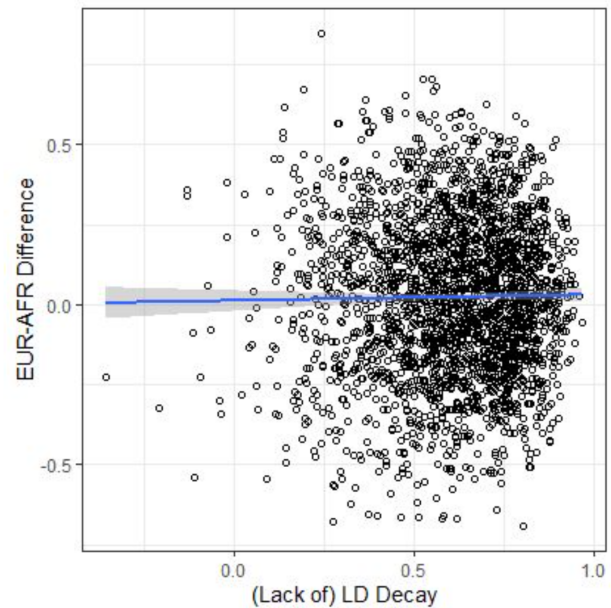

Figure 14: Association between LD decay and CEU-YRI difference in EA polygenic score.

ure 14), meaning that SNPs that are more robust to LD decay have similar allele frequency differences between Africans and Europeans.

However, after selecting the SNPs with low LD decay (r> 0.8), the difference increased from 2.32 % to 2.58 %. With even lower LD decay (r> 0.9), the difference became larger (3.46 %).

Similar results were obtained for Chinese (CHB) and CEU, although there was a non-significant trend towards SNPs with low LD decay to show larger PGS differences (r=0.022, p= 0.283. Figure 15). The overall amount of LD decay was smaller than for CEU-YRI (r= 0.728). After selecting the SNPs with low LD decay (r> 0.8), the difference increased from 1.36 % to 1.56 % but it increased to 3.8 % with even lower LD decay (r>0.9)

The same analysis was performed on the height significant SNPs. After filtering for MAF (>0.01) there were 2762 SNPs, and 1,693,394 variants in LD within the 500 Kb window.

The average correlation between the pairwise R2 values of CEU and YRI was r= 0.63. The polygenic score difference was 0.67 % (50.02 % vs 49.33 %). There was a significantly negative correlation between (lack of) LD decay and PGS difference (r= -0.0409, p= 0.0315). This can be seen from the scatterplot in Figure 16.

After selecting the SNPs with low amount of LD decay (r>0.8), the PGS difference was reduced to 0.53 %. With even lower LD decay (r>0.9), the difference became negative (-2.55 %), with YRI showing higher scores.
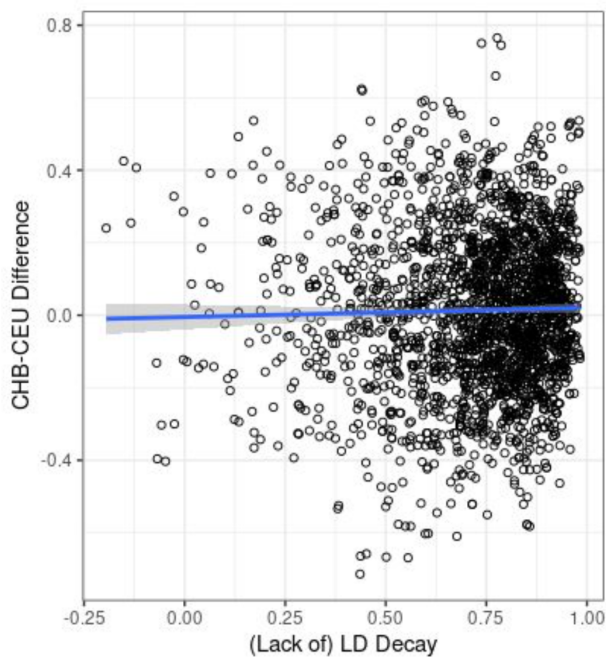
Figure 15: Association between LD decay and CHB-CEU difference in EA polygenic score.
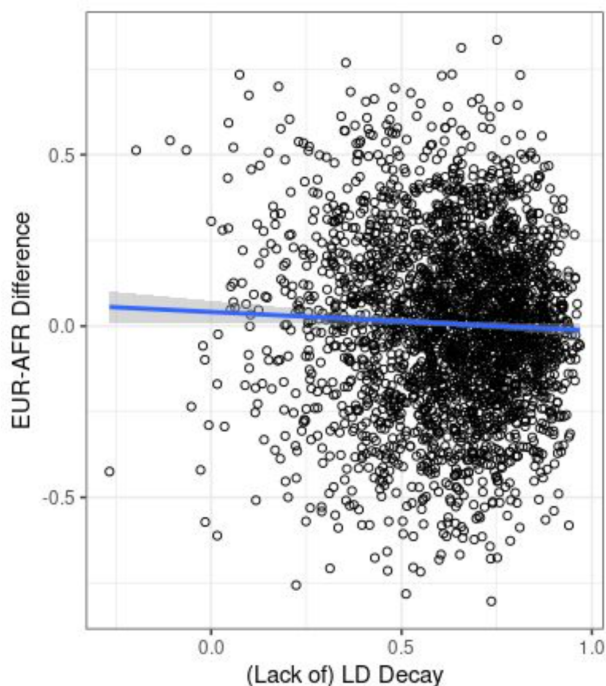

Figure 16: Association between LD decay and CEU-YRI difference in Height polygenic score.

A similar pattern was found for the CHB-CEU pair. There was a marginally significant negative correlation between (lack of) LD decay and PGS difference (r= -0.038, p= 0.055). The degree of LD decay was lower than for CEU-YRI (r=0.742) and the PGS difference was 1.17 %. This difference decreased to 0.84 % with LD > 0.8, and to 0.3 % with LD >0.9.

3.5 Cross-population LD decay

An interesting question is whether LD decays pattern change across populations: do the SNPs that are less subject to LD decay between two populations, tend to show lower LD decay between other pairs of populations?

The correlation between the LD decay across SNPs for CEU-YRI and CEU-CHB (using the EA SNPs) was 0.3, implying a weak homogeneity in LD decay patterns across different population pairs. This finding makes it difficult to select SNPs that are robust to LD decay across several ethnic groups. For example, there were only 234 SNPs that had low LD decay (r>0.8) in both the CEU-YRI and CEU-CHB pair.

However, it is possible to compute for each pair, the PGS deviation from the reference population after taking into account LD decay. To do this, the reference population (CEU) score is constrained to 0, and each population's score is represented by the deviation from 0. The deviation score can be calculated either using the SNPs in low LD (i.e. r>0.8) for each population, or by computing the polygenic score for each SNP weighted by its LD decay coefficient. Table 3 reports the LD-decay weighted PGS for 2 populations. Since computing LD decay requires long computation/server time, this analysis was limited to a few populations.

Table 3: LD-decay weighted PGS (CEU set to 0).

	CHB	YRI
EA3	0.0108	-0.0148
Height	-0.0928	-0.0024

3.6 Correlation between allele LD decay, Fst and allele frequency differences

Fst is a measure of allele frequency differences. However, unlike polygenic scores, it is non-directional, that is, it is "agnostic" about the allele's direction of effect. Hence, Fst suffers from loss of information compared to polygenic scores. Indeed, it was almost perfectly correlated (r= 0.97) to the absolute allele frequency difference (AFD, absolute value of the difference between population frequencies) (Figure 17), but had zero correlation with the polygenic score difference. Fst and AFD were also negatively correlated
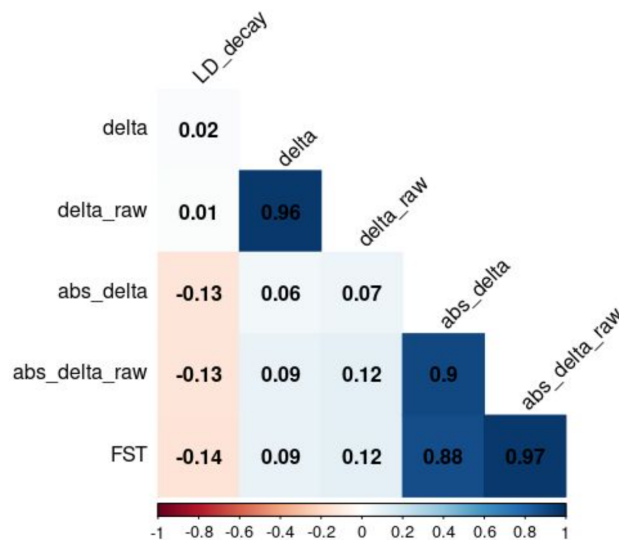
Figure 17: Correlation matrix (LD decay, PGS difference ("delta_raw", "delta") Fst ("FST") and absolute allele frequency difference ("abs_delta", "abs_delta_raw").

to (lack of) LD decay, that is SNPs with higher LD decay had higher Fst values.

The results were almost identical when using LD decay across European and East Asians (CEU - CHB). However, the extent of LD decay between CEU and CHB was lower than between CEU and YRI (r= 0.727 and 0.608, respectively; t= 22.043, p < 2.2e^{-16}).

3.7 PGS difference/Fst as signal of selection

If the polygenic score difference between populations is more sensitive to selection than Fst or AFD, the average ratio of PGS difference to Fst across SNPs should be higher for GWAS SNPs under selection than for less selected traits. Since Fst is nearly equivalent to the absolute allele frequency difference, it is a mixture of allele frequency shifts due to drift and to directional selection but it is not possible to disentangle them. Conversely, the PGS difference is more representative of selection pressure. Hence, the AFD/PGS difference ratio could be a useful indicator of selection pressure on a trait. The PGS difference, AFD and their ratios for European-Chinese and European-African pairs are reported in Table 4.

3.8 Simulation with quasi-random GWAS SNPs

The EA GWAS SNPs with p value >0.95 (N= 2331) and MAF> 0.01 were used to compute the PGS delta, the AFD and the correlation with LD Decay. A prediction of the polygenic selection model is that the PGS delta and the PGS delta/absolute delta will be lower than in the GWAS significant SNPs.

The polygenic scores were close to the theoretically expected value of 50 % (CEU= 0.4965 and YRI= 0.4997) and the difference was 0.3 %. The AFD (0.12) was almost 2 times lower than in the GWAS significant SNPs, and the raw PGS delta/AFD ratio was 4.5 times smaller (0.0259). The lower AFD (Fst) difference is due to the MAF being lower in the less GWAS significant SNPs (0.184 vs 0.311). Indeed, Fst is mathematically dependent on MAF, and has lower values with lower MAF (Jakobsson et al., 2013).

3.9 Difference in correlation coefficient between GWAS significant and quasi-random SNPs

Another metric to represent similarity in allele frequencies between a population pair is the correlation coefficient. Unlike the Fst, the correlation coefficient is sensitive to the sign of the difference between each pair of observations. In other words, it represents not only the strength of the relationship between two variables (or lack thereof) such as Fst or the absolute allele frequency difference; in fact, it is also an expression of the direction of such a relationship because it can acquire positive and negative values.

Therefore, directional (divergent) selection on a population pair should depress the correlation coefficient between the (same) alleles in the two populations. A prediction of this model is that the correlation between the allele frequencies of a pair of populations will be lower for GWAS significant SNPs than for less significant (or random) SNPs.

Indeed, the correlation between the YRI and CEU EA GWAS significant SNPs was r= 0.592 (95 % C.I.= 0.566/0.616), and r= 0.894 (95 % C.I.= 0.885/0.901) for the quasi-random SNPs.

The partial correlation of the population frequencies controlling for LD decay was nearly identical, indicating that LD decay does not mediate the correlation.

The height significant SNPs had similar correlation between CEU and YRI (r= 0.612) and CEU-CHS (r= 0.673).
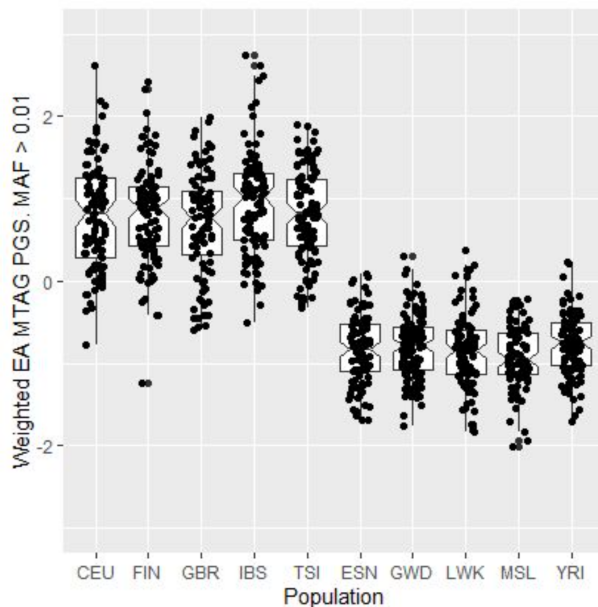
Overall, the correlations for the GWAS significant SNPs across populations were lower than for quasi-random SNPs, suggesting the presence of divergent selection at these loci.

3.10 Random SNPs

1000 sets of random SNPs, matched for LD score and MAF, were generated using SNPSNAP. To simulate a GWAS, the status of the effect allele was assigned at random for each SNP. The average correlation between the CEU and YRI allele frequencies over 1000 sets was 0.814.

Table 4: GWAS significant SNPs: PGS difference, AFD, PGS difference/AFD.

	Raw PGS delta	AFD	Raw PGS delta/AFD
CEU-CHB (EA3)	-0.0126	0.1759	0.0715
CEU-YRI (EA3)	0.0231	0.2026	0.1143
CEU - CHB (Height)	0.0148	0.1789	0.082
CEU-YRI (Height)	0.0067	0.2	0.0335

Figure 18: Individual EA3 polygenic scores by population.

3.11 *Individual PGS*

Polygenic scores were computed for each individual in 1000 Genomes for Europeans and Africans. The scores were standardized (Z-transformed), and 0 represents the mean of the entire sample.

The individual polygenic scores differed by 1.7 SDs across the two groups (t = -20.902, df = 152.74, p-value < 2.2e-16). The PGS were calculated for all the populations in the European and African 1000 Genomes groups. Most of the variation was between continents, with little overlap, but considerable overlap among populations from the same continental group (Figure 18).

It is possible to quantify the amount of variance within and between groups using ANOVA. One-way ANOVA was run with the PGS as the dependent variable, and population (CEU and YRI) as the independent variable. The F statistic represents the ratio of the variance between groups and within groups. A value higher than 1 implies that the variance between groups is larger than the variance within groups. Welch's one-way ANOVA was performed on CEU and YRI to account for different variance between groups. The F-statistic was 141.47 (p< 2.2 *10^{-16}).

4   Discussion

Height and education-associated SNPs were both highly differentiated across populations, as shown by the global Fst enrichment test and polygenic scores. These differences matched differences in average trait (i.e. height and education), reaching correlations 0.9 with average population IQ and height (Figures 3, 4, 6, 8, 9), implying that selection pressure after the out of Africa dispersal acted with different strength on different populations. The results were robust across different datasets (gnomAD, 1000 Genomes and HGDP), yielding more credibility to the findings. The lack of a correlation between LD decay and population differences in polygenic scores (Figure 14) suggests that low p value EA GWAS SNPs have a causal effect on the phenotype or that they are closely tagging causal variants (i.e. high LD). Another interpretation is that the tag variants represent random noise which does not bias the population means in one direction. However, if this was the case for all tag variants, the average allele frequency for both populations would be around 50 % (i.e. see the quasi-random SNPs), and there would not be a significant difference. Indeed, the SNPs with low LD decay (r> 0.9) had larger CEU-YRI frequency difference than the others (3.5 % vs 2.3 %), implying that the SNPs with most signal (and least noise) have larger frequency differences between populations.

For height, LD decay had a small but significant effect on the PGS difference, so that the PGS difference computed using SNPs in low LD decay (r> 0.8) was slightly reduced (from 0.67 % to 0.53 %). The SNPs with even lower LD decay (r> 0.9) produced a reversal of the difference, as Yorubans had higher PGS than CEU. This result must be interpreted with caution because it involves a small subset of SNPs. In this case, LD decay produces a small bias, inflating the European (or deflating the African) polygenic score.

Another way to interpret these findings is that the PGS differences between populations are driven by a subset of variants with very low LD decay (r> 0.8), whereas the majority of the variants add mostly noise which may or may not inflate the frequency differences. A detailed description of this phenomenon, however, is beyond the scope of this paper.

There was a lot of heterogeneity in LD decay patterns across populations (r= 0.3 for CEU-YRI and CEU-CHB with the EA GWAS), Hence, to increase polygenic score accuracy it is advisable to compute PGS differences selecting SNPs that have low LD decay for each specific pair of populations because the number of SNPs that are robust to LD decay across all populations is very small.

Climate is an environmental factor that might have influenced polygenic adaptation for education, via selection for enhanced cognitive ability or life-history traits. Indeed, a positive correlation (r= 0.68) between education PGS and latitude was found (Figure 5). A simulation using random SNPs matched for minor allele frequency, showed that this result would be highly unlikely under a model of random SNPs with similar minor allele frequency (p= 0.0009) (Figure 6). Remarkably, the height PGS had a much weaker correlation (r= 0.3) with latitude (Figure 1), despite Bergmann's rule predicting cold-climate selection for larger sized animals. However, this mirrors findings of a weak positive relationship between height and latitude in human populations (Gustafsson & Lindenfors, 2009).

Ethnic groups of East Asian origin had large positive residuals, that is to say their polygenic scores for EA were higher than predicted by the line of best fit when using latitude in a simple linear regression model. After adding the "sub-continent" variable to the regression model, there was a significant improvement in model fit (p= $2.64 * 10^{-7}$), with the variance explained (Adj. R2) increasing from 44.4 % to 75.8 %. On the other hand, the effect of latitude on EA PGS was more than halved (Beta= 0.67 to 0.28). This suggests the presence of selective pressures (or introgression events) independent of latitude and shared within sub-continents. An alternative explanation is that the current geographical location of some groups doesn't reflect their historical origin. For example, a positive residual for East Asian populations could be due to an origin at higher latitudes followed by a relatively recent move southwards.

Recently, Stern et al. (2021) found evidence for directional selection on EA3 PGS. The effect was partly mediated by EA3's correlation to a variable measuring skin pigmentation ("sunburning ability"). The authors found that after accounting for selection pressure on skin pigmentation, the selection signal on EA3 was attenuated. Hence, a large share of the selection pressure on EA3 was due to correlated response with another trait. In light of the strong correlation with latitude, an alternative explanation to directional selection on EA3 being a by-product of selection on pigmentation, is that climatic factors could account for the genetic correlation between EA3 and skin pigmentation. There are different mechanisms

that can potentially drive a correlated response to selection between two or more traits:

(a) passive, meaning they are genetically linked (i.e. pleiotropy) but only one trait is beneficial for passing on an organism's genes and the other traits "hitch-hike" because they are influenced by the same genes or genetic variants in strong linkage with the selected trait,

(b) the same environment selects for multiple traits at the same time, or

(c) different environmental selection pressures happen to act simultaneously on multiple traits.

The present study also shows the empirical equivalence of Fst and absolute allele frequency difference (r= 0.97). An advantage of using polygenic scores compared to the Fst or absolute allele frequency difference is that it is directional, that is, that each allele's effect is taken into account, whether it is a risk allele or not, or whether it increases or decreases a phenotype. On the contrary, Fst and AFD are non-directional, so the mean Fst or AFD across many genetic variants is independent of alleles with positive effects being overrepresented in a population compared to another one. Hence, Fst or AFD are more representative of drift than of selection. Conversely, the polygenic score difference is more indicative of directional selection because it is dependent on the average direction of selection across many genetic variants. Hence, the ratio between the latter and the mean Fst or AFD is a measure of directional selection net of the effects of drift. Accordingly, there was a much stronger reduction in PGS difference than Fst in a set of low significance GWAS SNPs (p>0.95) , which putatively contain mostly noise and very little selection signal. They had lower (2x) Fst, but much lower (7x) PGS difference, resulting in a 4.5-fold reduction in PGS difference/AFD ratio (Table 3).

Another way to represent the coefficient of directional selection is to compute the correlation between the frequencies of (GWAS significant) alleles with positive effect among a pair of populations; one can compare them to non-significant (almost random) SNPs, or a set of random SNPs matched for MAF. The advantage of using correlation of allele frequency across population pairs is that it is intuitive and, unlike Fst, sensitive to the direction of selection on each genetic variant as well as its strength. Divergent selection will make allele frequencies shift in different directions among populations. Indeed, the GWAS significant SNPs for height and EA had much weaker correlation of allele frequencies (r~ 0.6) compared to non-significant and random SNPs (r>0.8).

Allele frequencies computed by group are traditionally used for tests of selection. This allows researchers

to identify the genetic variants that have the strongest selection signals. However, another way to represent allele frequency differences between groups is with individual PGS: this gives a better idea of how much variation is partitioned between individuals and between groups. An in-depth analysis of this topic is not within the scope of this paper, but it can be seen in Figure 18 that for polygenic scores of traits under strong polygenic selection, there is much more dispersion between continental groups than between individuals. That is, there is little overlap in the individual polygenic scores across continental groups (Africans and Europeans), as they deviate by more than 1.5 SDs. On the other hand, populations within the same continent have relatively small differences. This suggests that genetic differences in cognitive abilities and life history are due to factors (cultural, climatic, or archaic introgression) that are fairly homogeneous within (sub-)continents. However, selection pressures within super populations do exist, as implied by the existence of specific groups with significantly higher scores compared to their reference super populations (e.g. Ashkenazi Jews, Finns) and the persistence of the effect of latitude after taking into account the super population variable.

Finally, the next generation of GWAS studies are increasingly looking at trans-ethnic samples to improve polygenic predictions in different populations. As GWAS on different populations are available, future studies should investigate whether the SNPs with lower LD decay found using European-based GWAS have higher predictive accuracy in non-European populations.

References

Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS genetics*, *10*(8), e1004412. doi: 10.1371/journal.pgen.1004412

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*(7), 1177–1186. doi: j.cell.2017.05.038

Chung, W., Chen, J., Turman, C., Lindstrom, S., Zhu, Z., Loh, P.-R., ... Liang, L. (2019). Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *Nature Communications*, *569*(2). doi: 10.1038/s41467-019-08535-0

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., ... Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, *9*(2098). doi: 10.1038/s41467-018-04362-x

Demange, P. A., Malanchini, M., Mallard, T. T., Biroli, P., Cox, S. R., Grotzinger, A. D., ... Nivard, M. G. (2021). Investigating the genetic architecture of noncognitive skills using gwas-by-subtraction. *Nature Communications*, *53*, 35–44. doi: 10.1038/s41588-020-00754-2

Eisenberg, D. T., Kuzawa, C. W., & Hayes, M. G. (2010). Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *American Journal of Physical Anthopology*, *143*, 13–20. doi: 10.1002/ajpa.21298

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton1, K. J., Golan, D., ... Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. *Science*, *354*(6313), 760–764. doi: 10.1126/science.aag0776

Fryar, C. D., Kruszon-Moran, D., Gu, Q., & Ogden, C. L. (2018). Mean body weight, height, waist circumference, and body mass index among adults: United states, 1999-2000 through 2015-2016. *National health statistics reports*, *122*, 1–16. doi: j.cell.2017.05.038

Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., ... Yang, J. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. *Nature Communications*, *9*(1865). doi: 10.1038/s41467-018-04191-y

Gustafsson, A., & Lindenfors, P. (2009). Latitudinal patterns in human stature and sexual stature dimorphism. *Annals of Human Biology*, *36*(1), 74-87. doi: 10.1080/03014460802570576

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., ... Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*, *24*, 169—181. doi: 10.1038/s41380-017-0001-5

Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., ... Kamatani, Y. (2020). Large-scale genome-wide association study in a japanese population identifies novel susceptibility loci across different diseases. *Nature Genetics*, *52*, 669—679. doi: 10.1038/s41588-020-0640-3

Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013). The Relationship Between FST and the Frequency of the Most Frequent Allele. *Genetics*, *193*(2), 515–528. doi: 10.1534/genetics.112.144758

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., ... Stefansson, K. (2017). Selection against variants in the genome associated

with educational attainment. *Proceedings of the National Academy of Sciences*, *114*(5), E727–E732. doi: 10.1073/pnas.1612113114

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, *50*(8), 1112–1121. doi: 10.1038/s41588-018-0147-3

Limborska, S. A., Balanovsky, O. P., Balanovskaya, E. V., Slominsky, P. A., Schadrina, M. I., Livshits, L. A., . . . Spitsyn, V. A. (2002). Analysis of CCR5Delta32 geographic distribution and its correlation with some climatic and geographic factors. *Human heredity*, *53*(1), 49—54. doi: 10.1159/000048605

Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., . . . Price, A. L. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics*, *47*, 1385—1392. doi: 10.1038/ng.3431

Pers, T. H., Timshel, P., & Hirschhorn, J. N. (2015). SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics*, *31*(3), 418–420. doi: 10.1093/bioinformatics/btu655

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*(2), 168—200. doi: 10.46469/mq.2013.54.2.3

Piffer, D. (2015). A review of intelligence gwas hits: Their relationship to country iq and the issue of spatial autocorrelation. *Intelligence*, *53*, 43-50. doi: 10.1016/j.intell.2015.08.008

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from gwas hits: A replication of previous findings using recent data. *Psych*, *1*(1), 55–75. doi: 10.3390/psych1010005

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS genetics*, *1*(6), e70. doi: 10.1371/journal.pgen.0010070

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., . . . Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics*, *50*(7), 912—919. doi: 10.1038/s41588-018-0152-6

Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. *American Journal of Human Genetics*, *99*(1), 139—153. doi: 10.1016/j.ajhg.2016.05.013

Stern, A. J., Speidel, L., Zaitlen, N. A., & Nielsen, R. (2021). Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. *American Journal of Human Genetics*, *108*(2), 219—239. doi: 10.1016/j.ajhg.2020.12.005

Uricchio, L. H., Kitano, H. C., Gusev, A., & Zaitlen, N. A. (2019). An evolutionary compass for detecting signals of polygenic selection and mutational bias. *Evolution Letters*, *3*(1), 69—79. doi: 10.1002/evl3.97

Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., van Eck, H. J., & van Eeuwijk, F. A. (2017). Evaluation of ld decay and various ld-decay estimators in simulated and snp-array data of tetraploid potato. *Theoretical and Applied Genetics*, *130*, 123—135. doi: 10.1007/s00122-016-2798-8

Whitlock, M. C., & Guillaume, F. (2009). Testing for Spatially Divergent Selection: Comparing QST to FST. *Genetics*, *183*(3), 1055-1063. doi: 10.1534/genetics.108.099812

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., . . . the GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry. *Human Molecular Genetics*, *27*(20), 3641-3649. doi: 10.1093/hmg/ddy271

Zeevi, D., Bloom, J. S., Sadhu, M. J., Ben Yehuda, A., Zangen, D., Levy-Lahad, E., & Kruglyak, L. (2019). Analysis of the genetic basis of height in large jewish nuclear families. *PLoS Genetics*, *15*(7), e1008082. doi: 10.1371/journal.pgen.1008082