A review of intelligence GWAS hits: their relationship to country IQ and the issue of spatial autocorrelation.

Davide Piffer

email: pifferdavide@gmail.com

Abstract

A review of published intelligence GWA studies was carried out.

The average frequency (polygenic score) of nine alleles positively associated with intelligence is strongly correlated to country IQ ($r = 0.91$). Factor analysis of allele frequencies produced a factor with a similar correlation to IQ (0.86). The majority of alleles (7/9) loaded positively on this factor. Allele frequencies varied by continent in a way that matched the average group-level of phenotypic intelligence. Average allele frequencies for intelligence GWAS hits had higher inter-population variability than background expectations or height GWAS hits. This may suggest stronger selection for intelligence than human height. Random sets of SNPs and F_{st} distances were employed to deal with the issue of spatial autocorrelation, due to population structure. GWAS hits were much stronger predictors of IQ than random SNPs. Regressing IQ on F_{st} distances did not significantly alter the results but it confirmed that, whilst evolutionarily neutral (genome-wide) genetic distances are indeed related to IQ differences between populations, the GWAS hits frequencies predict the latter above and beyond the former.

Introduction

Over the last few years, researchers have started moving away from the study of genetic evolution using a single-gene, Mendelian approach towards models that examine many genes together (polygenic). The more genes are involved in a given

phenotype, the more the signal of natural selection will be “diluted” across different genomic regions (because each gene accounts for a tiny effect) making it difficult to detect it using approaches focused on a single gene (Pritchard *et al.*, 2010; Piffer, 2014). A first attempt at empirically identifying polygenic selection was made by Turchin *et al.*, (2012) on two populations (Northern and Southern Europeans) and evidence for higher frequency of height increasing alleles (obtained from GWAS studies) among Northern Europeans was provided. A drawback of that study was the reliance on populations from a single continent and that crude pairwise comparisons (e.g. French vs. Italian) were used without correlating frequency differences to average population height. Moreover, the strength of selection was not determined.

Two different approaches to identify selection based on the correlation of allele frequencies across different populations have been recently developed by Piffer (2013) and Berg & Coop (2014).

Piffer’s method uses factor analysis of trait increasing alleles (found by GWA studies) as a tool for finding a factor that represent the strength of selection on a phenotype and the underlying genetic variation (Piffer, 2014a). An additional methodology consists of computing the correlation between genetic frequencies and the average phenotypes of different populations; then, the resulting correlation coefficients are correlated with the corresponding alleles’ genome-wide significance (p value). If the alleles contain selection signals, a positive correlation will be found, as alleles with high p value (more likely to be false positives) have a weaker correlation to average population phenotype (Piffer, 2014a). This is the “method of correlated vectors” (MCV) (Jensen, 1994).

To date, a few genes have replicated their association with intelligence. Rietveld *et al.* (2013)’s meta-analysis found ten SNPs that increased educational attainment, comprising three with nominal genome-wide significance and seven with suggestive significance. A recent study has replicated the positive effect of these top three SNPs (rs9320913, rs11584700 and rs4851266) on mathematics and reading performance in an independent sample of school children (Ward *et al.*, 2014). These SNPs were also

associated with g (general intelligence) in a sub-sample of Rietveld et al.'s original study.

Another SNP (rs236330), located within gene FNBP1L, showed a significant association with general intelligence, reported in two separate studies (Davies et al, 2011; Benyamin et al, 2013). This gene is strongly expressed in neurons, including hippocampal neurons and developing brains, where it regulates neuronal morphology (Davies et al, 2011).

Rietveld et al. (2014) using the proxy-phenotype method, found three SNPs significantly associated with cognitive performance in a sample overlapping with that used in their previous study (Rietveld et al., 2013).

More recently, a GWAS focusing on fluid intelligence found 13 genetic variants with genome-wide significance (Davies et al., 2015). However, only three SNPs were independent signals (not in LD with each other).

A significant association for the single-nucleotide polymorphism (SNP) rs17518584 (P -value = 3.28×10^{-9} after adjustment for age, gender and education) in an intron of the gene cell adhesion molecule 2 (CADM2) for performance on tasks of executive function and processing speed (Letter Digit Substitution Test/Digit Symbol Substitution Task) was found by a recent GWAS based on a fairly large sample ($N=5429-32070$) (Ibrahim-Verbaas et al., 2015).

The aim of this paper is to analyze the population allele frequency patterns of all the SNPs found to date (April 2015) to have a genome-wide significant association to intelligence or a related cognitive phenotype (general cognitive ability, processing speed) and test the hypothesis that they predict national cognitive ability above and beyond neutral population genetics mechanisms (i.e. population structure due to migration and drift) which is a type of spatial autocorrelation. The latter is a phenomenon that occurs when the values of variables sampled at nearby locations are not independent from each other (Tobler 1970). Usually this is a problem, because it violates the assumption of independently and identically distributed (i.i.d.) errors of most standard statistical procedures hence inflating type I errors.

The method proposed in this paper to account for population structure (i.e. spatial autocorrelation) is based on the correlation between F_{st} distances for the entire genome (or a random part of it) and distances (that is, the absolute number of the difference between any two populations) on the factor for all the populations. Two matrices representing genetic distances with N unique pair-wise comparisons are generated, where $N = n(n-1)/2$. Another matrix representing phenotypic distances (i.e. on average population IQ) is then created.

The test of the hypothesis that the factor does not merely represent population structure is articulated in two steps:

1. The correlation between the two matrices representing genetic distances is calculated. The lower it is, the more likely that the result is positive (that is, not due to population structure), as selection will skew distances away from background neutral variation due to random drift.
2. A regression of phenotypic distances on factor distances + genome-wide F_{st} distances is carried out. If factor distances have an independent positive effect on the dependent variable (phenotypic distances), then the result is more likely to be genuine.

Methods

The factor extracted from the 4 replicated SNPs(henceforth “4 SNPs g factor”) was obtained by Piffer (2015a).

Allele frequencies for the SNPs were downloaded from 1000 Genomes, using the final release of phase 3 data: <http://browser.1000genomes.org/index.html>

IQs were obtained from Lynn & Vanhanen (2012). Finland’s and Vietnam’s IQ were adjusted upwards to 101 (from 97 in Lynn & Vanhanen), to account for recent, more accurate estimates (Armstrong et al., 2014) and (Rindermann et al., 2013). IQ for Tuscany was calculated as the average between the IQ estimated from PISA Creative

Problem Solving (Piffer & Lynn, 2014b) and from PISA Math, Science, Reading. There were 3 missing cases (Chinese Dai, Gujarati Indian, Indian Telegu) for which IQ was unknown.

Results

The correlation between the most recent GWAS hit (allele T of rs17518584) and the 4 SNPs g factor was $r=0.96$ ($N=26$).

The two hits (rs17522122.G, rs10119.G) from Davies et al. (2015) that were not in LD with any of the 4 replicated SNPs (thus excluding rs10554471, in LD with rs9320913 from Rietveld et al., 2013) were correlated to the 4 SNPs g factor. Pearson's r was respectively= -0.514 and 0.698.

One hit (rs1487441) located on chromosome 6 (pos. 98660615) from Rietveld et al. (2014) was in LD with one hit from Davies et al.(rs10554471) and from Rietveld et al., 2013 (rs9320913). The other two hits were independent signals (rs7923609 and rs2721173). Their correlation with the 4 SNPs g factor was, respectively $r= 0.322$ and - 0.697).

A factor analysis was then carried out on all the 9 SNPs (1 from Ibrahim-Verbahas et al., 2015, 2 from Davies et al., 2015, 2 from Rietveld et al., 2014, plus the 4 SNPs analyzed in Piffer, 2015a).

A single factor was extracted that explained 61% of the variance.

Structure matrix is shown in table 1.

Table 1. Structure matrix for 9 g increasing alleles.

Chr. #	Location	SNP	Factor loading
6*	98572120	rs10457441.C	0.70

14	32372633	rs17522122.G	-0.52
19	50098513	rs10119.G	0.78
1	204576983	rs11584700.G	0.82
2	100818479	rs4851266.T	0.93
1	94059554	rs236330.C	0.91
3	85604923	rs17518584.T	0.97
10	65133822	rs7923609.G	0.35
8	145744429	rs2721173.C	-0.84

*In LD with two hits from two published GWA studies: rs9320913, location: 98,584,733 (Rietveld et al., 2013); rs1487441, location: 98,553,894 (Rietveld et al., 2014). Another recent study (Trampush et al., 2015) replicated the effect of this locus (specifically, rs1906252) on cognitive function.

Factor scores are shown in table 2.

Table 2. Factor scores for 9 g increasing alleles.

Continents*	Population	Polygenic Score	9 SNPs g factor scores	4 SNPs g factor scores	IQ

AFR	Afr.Car.Barba dos	0.3726	-1.3174	-1.26112	83
AFR	US Blacks	0.3909	-1.2251	-1.21019	85
AFR	Esan Nigeria	0.3607	-1.6754	-1.45081	71
AFR	Gambian	0.3451	-1.5180	-1.44724	62
AFR	Luhya Kenya	0.334	-1.6720	-1.5391	74
AFR	Mende Sierra Leo	0.3516	-1.4886	-1.2412	64
AFR	Yoruba	0.3391	-1.6197	-1.4649	71
HISP	Colombian	0.4852	0.1390	-0.12223	83.5
HISP	Mexican LA	0.4871	0.3748	0.02157	88
HISP	Peruvian	0.5006	0.3496	-0.30414	85
HISP	Puerto Rican	0.4792	0.1013	0.00753	83.5
E.ASN	Chinese Dai	0.5568	1.2361	1.18278	N/A
E.ASN	Han Chinese Beijing	0.6182	1.2349	1.39839	105
E.ASN	Han Chinese South	0.6	1.1606	1.30377	105
E.ASN	Japanese	0.6076	0.9399	1.2297	105

E.ASN	Vietnam	0.5914	1.2287	1.59826	99.4
EUR	Utah Whites	0.5298	0.4879	0.75587	99
EUR	Finns	0.54	0.5797	0.71432	101
EUR	British	0.5427	0.5357	0.84863	100
EUR	Spanish	0.5294	0.3580	0.59903	97
EUR	Tuscan Italy	0.5229	0.4469	0.56805	99
SAS	Bengali Banglad.	0.4858	0.1920	-0.25727	81
SAS	Gujarati Ind. Tx	0.5126	0.5075	0.47096	N/A
SAS	Indian Telegu UK	0.5066	0.2838	-0.60945	N/A
SAS	Punjabi Pakistan	0.4976	0.2230	0.18886	84
SAS	Sri Lankan UK	0.4754	0.1371	-0.60945	79

*AFR= Sub-Saharan African; HISP= Hispanic/Latin American; E.ASN= East Asian; Eur= European; SAS= South Asian

There was a positive correlation between the 9 SNPs g factor and IQ ($r=0.863$, $N=23$).

The method of correlated vectors was used to assess the predictive validity of factor analysis. The SNP's correlation with national IQs was correlated to their factor loadings.

There was a positive correlation between the two variables ($r=0.986$).

Their relationship is plotted in figure 1.

Figure 1: MCV of factor loadings and correlation with national IQ.

Randomization

40 random SNPs matched to the 9 GWAS hits were obtained using SNPSNAP (<http://www.broadinstitute.org/mpg/snpssnap/>). These were used to test the hypothesis that the signal provided by the GWAS hits is distinguishable from background noise.

That is to say, the factor extracted from the GWAS hits should have better predictive power of national IQ than randomly matched SNPs.

10 sets of 4 and 4 sets of 9 SNPs were factor analyzed and the resulting factors were entered in a regression of IQ on the 4 and 9 SNPs g factor, respectively. A polygenic score will also be entered in the regression as predictor along with polygenic scores obtained from random SNPs. This was calculated as the average frequency of the 9 g increasing alleles. Its correlation to IQ was $r=0.91$ ($N=23$). The relationship is plotted in figure 2. Beta coefficients are reported in tables 3,4,5.

Figure 2. Relationship between national IQ and polygenic score.

ACB= African Caribbean in Barbados; ASW= Americans of African Ancestry in SW USA; BEB= Bengali from Bangladesh; CDX= Chinese Dai in Xishuangbanna, China; CEU= Utah Residents with Northern and Western European Ancestry; CHB= Han Chinese in Beijing, 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= Gambians in Western Divisions in the Gambia; Iberian Population in Spain; ITU= Indian Telegu from the UK; 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; PJI= Punjabi from Lahore, Pakistan; PUR= Puerto Ricans from Puerto Rico; STU= Sri Lankan Tamil from the UK; Toscani in Italia; Yoruban in Ibadan, Nigeria.

Table 3. Regression of 4 SNPs g factor and Random SNPs factors on IQ. Beta coefficients.

	Random SNPs Factor	4 SNPs g factor (GWAS hits)
1	-0.058	0.867
2	0.130	0.993
3	-0.29	0.655
4	-0.087	0.866
5	-0.092	0.991
6	0.081	0.985

7	0.038	0.946
8	-0.155	1.051
9	-0.172	0.759
10	0.017	0.930
Average Beta	0.112	0.9043

Table 4. Regression of 9 SNPs g factor and Random SNPs factors on IQ. Beta coefficients.

	Random SNPs Factor	9 SNPs g factor (GWAS hits)
1	-0.372	0.569
2	0.014	0.849
3	0.765	1.611
4	0.869	1.695
Average Beta	0.505	1.181

Table 5. Regression of GWAS hits polygenic score and random polygenic scores on IQ. Beta coefficients.

	Random SNPs Polygenic Score	9 GWAS hits Polygenic Score
--	-----------------------------	-----------------------------

1	-0.049	0.875
2	-0.180	0.964
3	0.487	1.344
4	-0.169	0.825
Average Beta	0.221	1.002

The average Beta was calculated using the absolute value for the random SNPs and the real number for the GWAS hits. This inflated the values of the random SNPs Betas but it is based on the conservative scenario that the majority of GWAS hits factor loadings are positive only by chance. GWAS hits produced higher Betas (1.03) than the random SNPs (0.279).

For the analyses using 4 and 9 SNPs, the GWAS hits factor was a better predictor of IQ than the random SNPs factor in all the 14 instances. For the analyses using the polygenic scores, the GWAS hits were a better predictor in 4/4 instances.

MCV was run on the four random sets of 9 SNPs. Contrary to expectations, it produced high correlations. These are reported in table 6.

Table 6. MCV of IQ and factor loadings for sets of 9 random matched SNPs.

First Set	Second Set	Third Set	Fourth Set
-0.9262672007	0.9570990262	-0.9780350457	-0.9739700128

This result is probably due to spatial autocorrelation. In this particular instance, spatial autocorrelation is due to populations closer in space also being genetically more similar.

This accounts also for another phenomenon: the correlations between factors extracted from the random sets of SNPs and IQ or the GWAS hits factors tend to be high, as shown in the correlation matrix reported in the supplementary files.

Although in the regression analyses the GWAS hits factors are better predictors of IQ than the random factors in the majority of cases, the latter still show a strong correlation to IQ. The average correlation with IQ (absolute correlation coefficient) is $r=0.74$. The average correlation between the GWAS hits factors and IQ is $r=0.89$.

It is thus necessary to deal with spatial autocorrelation by controlling for genome-wide genetic distances, with the procedure employed by Piffer (2015b).

Controlling for population structure

The same procedure applied by Piffer (2015b) will be extended to the 9 GWAS hits factor and Fst distances (Weir & Cockerham, 1984) for Chromosome 21 and 1 (largest and smallest) will be used. These were calculated using Vcftools on the 1000 Genomes, phase 3 files, downloaded from

<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>

Vcftools and R code is reported in the appendix.

In addition, distances calculated for the fourteen random SNPs factors will be used in a simulation: they will be entered as independent variable in a regression of IQ on Fst distances, instead of the GWAS hits factor. If they fail to predict IQ above and beyond Fst distances, but the GWAS hits factors do, then this will suggest that the GWAS hits factor has genuine predictive power independent of population structure.

Inspection of the correlation matrix (<https://docs.google.com/spreadsheets/d/1EanD-Jj15a3OBJpu5Tr5e86iLTLZPpa9vUEPTwaVsHU/edit?usp=sharing>) reveals that IQ distances have a stronger correlation to the GWAS hits distances ($r \times gdist = 0.78$; $gdist_9 = 0.65$; Polygenic = 0.76) than to the distances representing population

structure($r \times F_{st} \text{ Chr.1}=0.58$; $F_{st} \text{ Chr.21}= 0.58$; randomnine1=0.52; randomnine2=0.64; randomnine3=0.57; randomnine4=0.52).

A regression analysis was used to estimate the relative strength of each predictor.

Regression analysis

A total of 325 pairwise comparisons was obtained for the 26 populations of 1000 Genomes. Distances were calculated as the absolute value of the difference between population pairs on the selected variables (IQ, GWAS hit factors, polygenic score). The IQ variable had missing values, so a total of 253 distances was calculated. F_{st} distances were calculated for Chromosome 1, using the methodology employed by Piffer (2015b). Chr 1 and 21 F_{st} distances were almost identical ($r=0.992$), hence only the bigger chromosome was used. Beta coefficients are reported in tables 8a/b

Table 7a. GWAS hits distances and F_{st} : Standardized Beta coefficients.

Dependent variable: IQ.

Fst distances	
-0.093	4 SNPs g factor distances: 0.850
-0.033	9 SNPs g factor distances: 0.685
-0.409	Polygenic score distances: 1.227

Table 7b. Random SNPs factor distances and F_{st} : Standardized Beta coefficients.

Dependent variable: IQ.

Fst distances	
0.431	Random factor 1 distances: 0.317
0.036	Random factor 2 distances: 0.609
0.372	Random factor 3 distances: 0.227
0.626	Random factor 4 distances: -0.049

The ratio of average Betas (Average Beta random factor/Average Beta Fst) is $0.3/0.366 = 0.82$.

The ratio of Betas of GWAS hits to Fst Betas is $0.885/0.167 = 5.3$.

ANOVA

The average frequencies of the 9 GWAS hits for the five 1000 Genomes continental groups were calculated. These are represented in a boxplot (figure 3).

Figure 3. Average frequency of g increasing alleles by continental group.

ANOVA was conducted to analyze the difference between group means. $F(1.235)$, $\text{Pr}(> F) = 0.311$.

Tukey's post-hoc test was used to compare means. Confidence intervals are reported in table 9.

Table 8. Tukey's test with 95% confidence intervals for difference between continental group means.

	Difference	95% C.I.
--	------------	----------

AMR-AFR	0.103	-0.217, 0.424
ASN-AFR	0.237	-0.083, 0.557
EUR-AFR	0.174	-0.146, 0.494
SAS-AFR	0.140	-0.180, 0.461
ASN-AMR	0.133	-0.186, 0.454
EUR-AMR	0.070	-0.249, 0.391
SAS-AMR	0.037	-0.283, 0.357
EUR-ASN	-0.062	-0.383, 0.257
SAS-ASN	-0.097	-0.417, 0.224
SAS-EUR	-0.034	-0.354, 0.287

Estimating inter-population variability from average allele frequencies.

Inter-population variability is usually used as a way to detect signals of selection at specific loci. The F_{st} index is measured at a single locus, as it compares inter-population variability to within-population variability. Deviation from normality (the average genome-wide F_{st} value between two populations) suggest the presence of selection at that locus. Another approach should be applied to polygenic traits and it should be based on analyzing many loci together. Once the average allele frequency of trait-increasing alleles is calculated, it is possible to obtain simple measures of inter-population variability, such as the standard deviation (SD). The SD of average allele frequency of the 9 GWAS hits was 0.088. This was higher than the SD for the average frequency of sets of 9 SNPs: 0.043; 0.032; 0.078; 0.031.

Seven sets of 9 SNPs were also built from the 66 height GWAS hits reported in Piffer (2015c). The average SD was 0.06 (0.073, 0.057, 0.0658, 0.0751, 0.0286, 0.0607, 0.0587).

Discussion

The average frequency (polygenic score) of nine alleles positively associated with intelligence in published GWAS is strongly correlated to country IQ ($r = 0.91$). Factor analysis of allele frequencies produced a factor with a similar correlation to IQ (0.86). The majority of alleles (7/9) loaded positively on this factor. A set of 40 unrelated SNPs was drawn at random and their frequencies factor analyzed to calibrate background expectations. The pattern of very high positive or negative correlations suggested that spatial autocorrelation inflated the relationships between variables. That is to say, factors extracted from random SNPs showed very high correlations to the GWAS hits factors ($r = 0.6-0.98$) and similarly high correlation to IQ distances.

Unexpectedly, the method of correlated vectors produced very high values also when ran on the random SNPs, making the extremely high correlation (0.99) found for the GWAS hits less impressive. However, the GWAS hit factors tend to have a higher correlation to IQ than the random factors (0.89 vs. 0.74). When entered in a regression, the GWAS hits factors and the polygenic score were a better predictor of IQ than the random SNPs factor in all the instances (Average Betas of 1.03 and 0.28, respectively).

Additionally, F_{st} distances were calculated in order to provide an estimate of genetic distances between populations that reflect neutral evolutionary processes (e.g. migration and drift) but not selection pressure. Then, distances between populations were calculated as the absolute number of the difference between pairs (325 unique pairs). First, the 4 and 9 GWAS hits factors and the polygenic score emerged as better predictors of IQ than F_{st} distances (Betas between 0.68-1.12 vs 0-0.41). Paradoxically, F_{st} distances had a negative relationship with IQ when regressed with the polygenic score. The latter had a very high predictive power (Beta=1.12). A set of regressions was run with F_{st} but with random factors instead of GWAS hits factors as predictors. This showed that random factors did not predict IQ better than F_{st} distances. More precisely, GWAS hits had standardized beta coefficients about 5 times higher than the random

SNPs. Conversely, the latter had betas equal to or lower than F_{st} distances (ratio= 0.82).

Comparison of allele frequency means for the five continental groups from 1000 Genomes showed differences that closely correspond to estimated continent-level intelligence, with the following hierarchy: East Asian; European; South Asian; American (Hispanic); African. However, ANOVA did not result in p values that meet the conventional significance threshold ($p < 0.05$) and Tukey's test produced loose confidence intervals. Obviously, the lack of statistical significance is due to the very small sample size ($N=9$). The increasing number of GWA studies will undoubtedly provide more hits in the near future and it will be possible to have a more accurate picture of cognitive-related genetic variation within and between populations.

When comparing the average frequencies of height and intelligence GWAS hits, a striking difference is the much higher inter-population variability of the latter, ranging from 33.4% (Luhya, Kenya) to 61.82% (Han Chinese, Beijing). Conversely, the average frequencies of height increasing alleles range from 44.76% (Vietnam) to 49.5% (Esan Nigeria) (Piffer, 2015c). This could indicate that selection for intelligence has been stronger than selection for height, thus creating larger variability across isolated populations. The higher number of SNPs comprised by the height polygenic score ($N=66$) also likely accounts for the smaller differences but subsets of 9 height increasing SNPs had a lower average SD (0.06 vs 0.088). Moreover, the polygenic scores comprising 9 random SNPs have smaller average inter-population variability (SD= 0.046 vs 0.088). Future studies should use more sets of random SNPs to calibrate background expectation of inter-population variability of allele frequencies.

References:

Armstrong, E.L., Woodley, M.A., Lynn, R. Cognitive abilities amongst the Sàmi population. *Intelligence*, 2014: 35-39. doi: <http://doi.org/10.1016/j.intell.2014.03.009>

Benyamin, B., Pourcain, B.St., Davis, O.S., Davies, G., Hansell, N.K., Brion, M.-J.A. et al., Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Molecular Psychiatry*, 2013: 19: 253-258. <http://doi.org/doi:10.1038/mp.2012.184>

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S.E., Liewald, D., Xiayi, K., Le Hellard, S. et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, 2011: 996-1005. doi: <http://doi.org/10.1038/mp.2011.85>

Davies, G., Armstrong, N.,... Deary, I.J. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Molecular Psychiatry*, 1-10.doi:10.1038/mp.2014.188

Dormann, C.F., McPherson, J.M., ...and Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30: 609-628.

Ibrahim-Verbaas, C.A., Bressler, J.,Mosley, T.H. (2015). GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Molecular Psychiatry*, 1-9. doi:10.1038/mp.2015.37

Jensen, A. R., and Weng, L. J. (1994). What is a good g? *Intelligence* 18, 231-258.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Piffer, D. (2014a). Simple statistical tools to detect signals of recent polygenic selection. *IBC*, 6: 1-6. DOI: 10.4051 / ibc.2014.6.1.0001

Piffer, D. & Lynn, R. (2014b). New evidence for differences in fluid intelligence between north and south Italy and against school resources as an explanation for the north-south IQ differential. *Intelligence*, 246-249. doi: <http://doi.org/10.1016/j.intell.2014.07.006>

Piffer, D. (2015a). Estimating the genotypic intelligence of populations and assessing the impact of socioeconomic factors and migrations. *The Winnower* 2:e142299.93508. DOI:10.15200/winn.142299.93508

Piffer, D. (2015b), Intelligence GWAS hits: Selection signal or population structure? A test of the null hypothesis, *The Winnower*:e142593.30710
DOI:[10.15200/winn.142593.30710](https://doi.org/10.15200/winn.142593.30710)

Piffer D. (2015c). Evidence of polygenic selection on human stature inferred from spatial distribution of allele frequencies [v1; ref status: awaiting peer review, <http://f1000r.es/4ye>] *F1000Research*, 4:15 (doi:10.12688/f1000research.6002.1)

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H. J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 1467-1471. DOI: <http://doi.org/10.1126/science.1235488>

Rietveld, C.A., Esko, T., ... and Koellinger, P.D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *PNAS*, 111: 13790-13794. doi: 10.1073/pnas.1404623111

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. ! *Econ. Geogr.* 46: 234-240.

Trampush, J.W., Lencz, T.,...Malhotra, A.K. (2015). Independent evidence for an association between general cognitive ability and a genetic locus for educational attainment. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, doi: 10.1002/ajmg.b.32319

Turchin, M. C., Chiang, C. W., Palmer, C. D., Sankararaman, S., Reich, D., Genetic Investigation of, A. T. C., and Hirschhorn, J. N. (2012). Evidence of widespread

selection on standing variation in Europe at height-associated SNPs. *Nat Genet* 44, 1015-1019.

Ward, M.E., McMahon, G., St Pourcain, B., Evans, D.M., Rietveld, C.A., et al. Genetic Variation Associated with Differential Educational Attainment in Adults Has Anticipated Associations with School Performance in Children. *PLoS ONE*, 2014: 9, e100248.
doi:<http://doi.org/10.1371/journal.pone.0100248>

Weir, B.S., & Cockerham, C.C. Estimating F-Statistics for the analysis of population structure. *Evolution*, 38: 1358-1370. (1984)

Appendix:

Vcftools code

```
cd c:/folder/... #set to directory containing 1000 Genomes vcf file
```

```
c:/Users/Davide/vcftools/bin/vcftools --
```

```
vcfALL.chr1.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf --weir-  
fst-pop POP1.txt --weir-fst-pop POP2.txt --out fst.POP1.POP2
```

R code and datasets can be downloaded from: osf.io/jt73x