



Intelligence GWAS hits: Selection signal or population structure? A test of the null hypothesis

DAVIDE PIFFER

ABSTRACT

An index of population structure (F_{st}) is used to test the null hypothesis that the genetic factor extracted from GWAS hits represents differences between populations due to migrations and drift. Employing the 1000 Genomes data, a regression of average IQ distances on the general intelligence genetic factor and F_{st} distances shows that the former is the only significant predictor of IQ distances ($\text{Beta} = 0.82$), whereas the population structure has no independent predictive power ($\text{Beta} = -0.05$). This result suggests that the null hypothesis can be rejected.

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

pifferdave@gmail.com

DATE RECEIVED:

June 10, 2015

DOI:

10.15200/winn.142593.30710

ARCHIVED:

March 09, 2015

KEYWORDS:

population structure

CITATION:

Davide Piffer, Intelligence GWAS hits: Selection signal or population structure? A test of the null hypothesis, *The Winnower* 2:e142593.30710, 2015, DOI: [10.15200/winn.142593.30710](https://doi.org/10.15200/winn.142593.30710)

© Piffer This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



INTRODUCTION

Piffer(2015) reported a factor purportedly indicating the strength of selection on intelligence for populations of the 1000 Genomes database. A test of the null hypothesis that this result was due to population structure or drift was provided. However, it was fairly primitive for two reasons: 1) it relied on continental-level data and not populations, thus dramatically reducing the resolution; 2) Even worse, it relied on an indirect estimate of genome-wide distances based on ancestral component published on a blog.

The most commonly used measure of population differentiation is the Fixation Index (F_{st}). This represents the average population differentiation at a given locus or across the entire chromosome or genome.

The method proposed in this paper 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 AND RESULTS

The genotypic “intelligence” factor (henceforth “g factor”) reported by Piffer (2015) was used. Fst distances were calculated using Vcftools (<http://vcftools.sourceforge.net/>) v0.1.13, which is based on Weir and Cockerham (1984).

The variant set was downloaded from the 1000 Genomes, using the final release of phase 3 data: <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>

Vcftools and R code is reported in the Appendix. The dataset can be downloaded from this url: <https://docs.google.com/spreadsheets/d/1evV85pJyLaBXEShV3WUk6PbTB7kt73csOiY8iZP43ms/edit?usp=sharing>

A total of 325 pair-wise comparisons were generated from the 26 populations (26*25/2). Fst distances for chromosome 21 were calculated using Vcftools. This chromosome was chosen as it is the smallest hence requiring less CPU time. Distances on the g factor were calculated as the absolute number of the difference on the factor score between the 26 populations. There was a significant correlation between genome-wide distances and g factor distances ($r = 0.785$, $N=325$, $p<0.0001$). Correlations between g factor, IQ and Fst distances are reported in table 1

Table 1: Correlation matrix

	IQ Distances	Fst Distances (Chr.21)
IQ Distances		
Fst Distances (Chr.21)	0.588	
G factor distances	0.776	0.786

A multiple linear regression was carried out of IQ distances on Fst and 4 SNPs g factor distances, resulting in 253 cases after list-wise deletion of missing data (NA= 72) Only the latter emerged as a significant predictor (Beta= 0.82), whilst the former did not (Beta= -0.056).

DISCUSSION

This paper illustrates a way to test Piffer’s factor analytic method against the null hypothesis that the factors merely represent population structure and not a selective process.

The factor extracted from the SNPs associated with intelligence within population via GWAS, appear to bear a genuine signal of recent selection that predicts population differences in IQ above and beyond genome-wide distances which reflect drift and admixture due to migration between human populations.

This method can be fruitfully applied to other analyses involving selection differentials between populations.

REFERENCES

Piffer, D. Estimating the genotypic intelligence of populations and assessing the impact of socioeconomic factors and migrations. *The Winnower* 2:e142299.93508 (2015).
DOI:10.15200/winn.142299.9350

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 --vcf
ALL.chr21.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf --weir-fst-pop POP1.txt --
weir-fst-pop POP2.txt --out fst.POP1.POP2
```

R code

```
BetaFst <- read.csv("~/fstproject/BetaFst.csv")

View(BetaFst)

setwd("~/fstproject")

IQdistances=BetaFst$IQ.distances

fst_21=BetaFst$Chr21.Fst.1

gfdistances=BetaFst$X4.SNPs.GI.distances

library(QuantPsys)#required for lm.beta

df=data.frame(IQdistances,fst_21,gfdistances)

newdata <- na.omit(df)#delete missing values

cor(newdata)#correlation matrix

fit <- lm(IQdistances ~ fst_21 + gfdistances, data=newdata)#multiple linear regression
```

beta=lm.beta(fit)#standardized beta coefficients

structure(beta)