

Estimating the genotypic intelligence of populations and assessing the impact of socioeconomic factors and migrations.

Davide Piffer¹

¹Affiliation not available

April 17, 2023

Abstract

Factor analysis of allele frequencies was used to identify signals of polygenic selection on human intelligence. Four SNPs which reached genome-wide significance in previous meta-analyses were used. Allele frequencies for 26 population were obtained from 1000 Genomes. The resulting factor scores were highly correlated to average national IQ ($r=0.92$). A regression of IQ differences between subcontinental groups on the 4 SNPs g factor and an index of genome-wide genetic distances showed the former was an independent and significant predictor (Beta= 1.14), whereas genome-wide distances lost all predictive power. This finding suggests that the relationship between the 4 SNPs g factor and IQ is due to natural selection on a specific phenotype and not the result of a spurious correlation arising from genome-wide evolutionary processes such as random drift or migrations. A regression of IQs on genetic factor scores of developed countries was used to estimate the predicted genotypic IQs of developing countries. The residuals (difference between predicted and actual scores) were negatively correlated to per capita GDP and Human Development Index, implying that countries with low socioeconomic conditions have not yet reached their full intellectual potential.



Estimating the genotypic intelligence of populations and assessing the impact of socioeconomic factors and migrations.

DAVIDE PIFFER

ABSTRACT

Factor analysis of allele frequencies was used to identify signals of polygenic selection on human intelligence. Four SNPs which reached genome-wide significance in previous meta-analyses were used. Allele frequencies for 26 populations were obtained from 1000 Genomes. The resulting factor scores were highly correlated to average national IQ ($r=0.92$). A regression of IQ differences between subcontinental groups on the 4 SNPs g factor and an index of genome-wide genetic distances showed the former was an independent and significant predictor ($\text{Beta} = 1.14$), whereas genome-wide distances lost all predictive power. This finding suggests that the relationship between the 4 SNPs g factor and IQ is due to natural selection on a specific phenotype and not the result of a spurious correlation arising from genome-wide evolutionary processes such as random drift or migrations. A regression of IQs on genetic factor scores of developed countries was used to estimate the predicted genotypic IQs of developing countries. The residuals (difference between predicted and actual scores) were negatively correlated to per capita GDP and Human Development Index, implying that countries with low socioeconomic conditions have not yet reached their full intellectual potential.

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:
pifferdavide@gmail.com

DATE RECEIVED:
June 11, 2015

DOI:
10.15200/winn.142299.93508

ARCHIVED:
February 03, 2015

KEYWORDS:
psychology, intelligence,
evolutionary genetics,
polygenic selection

CITATION:
Davide Piffer, 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.142299.93508

INTRODUCTION

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 FBNP1L, 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).

Piffer (2013) applied principal components analysis (PCA) to allele frequencies to obtain an estimate of natural selection (or deviation from random drift) on different alleles correlated to the same phenotype.

The aim of this paper is to provide updated genotypic IQ scores for populations, by using the updated 1000 Genomes database (phase 3) comprising 26 populations instead of 14 and different factor analytic methods instead of PCA.

Another aim of this paper is to test the hypothesis that a detrimental environment can depress average phenotypic IQ, hence populations living in worse socioeconomic conditions would not have reached

© 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.



their full potential (as indicated by their genotypic score).

METHODS

IQ/educational attainment increasing alleles: Rietveld et al's top 3 hits were included together with another SNP(rs236330), located within gene FBNP1L, reported in two separate studies (Davies et al, 2011; Benyamin et al, 2013).

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, 2014) and from PISA Math, Science, Reading. There were 3 missing cases (Chinese Dai, Gujarati Indian, Indian Telegu).

Two socio-economic indicators were used: Gross Domestic Product (World Bank, 2014) at purchasing power per capita (GDP (PPP)) and the Human Development Index for 2014 (United Nations Development Programme, 2014).

IQ for Tuscany was obtained from Piffer & Lynn (2014) as the IQ of Central Italy. IQ for Vietnam was obtained from Rindermann et al. (2013).

Genome-wide distances (Fst) were obtained from Gedmatch K13 (2013).

Piffer (2013) used Principal Components Analysis of population data from 1000 Genomes, phase 1, which had data for only 14 populations from four racial clusters and from 50 populations contained in ALFRED. Besides ALFRED, here I use the last updated 1000 Genomes phase 3 data, comprising 26 populations from five continental groups. I employed factor analysis instead of principal components analysis because it is the preferred method when the purpose is identifying a latent structure free from unique variance (i.e. error), which in the case of allele frequencies can be due to random drift (shifting frequencies randomly up or down) or inadequate (i.e. small) sampling.

RESULTS

FACTOR ANALYSIS OF 4 IQ INCREASING ALLELES.

A previous study found that the specific factor extraction method employed did not affect results much except for the use of principal components analysis which produced inflated loadings (Kirkegaard, 2014). To further examine how factor extraction method influences results several methods were employed (minimum residuals, weighted least squares, generalized least squares, principal axis factoring, maximum likelihood) and factor scores were obtained using different methods (Thurstone, Harman, and Bartlett). These all produced nearly identical results, yet they were averaged to create a composite vector. The composite factor had slightly higher validity, as suggested by its slightly higher correlation with Lynn and Vanhanen's national IQs ($r=0.92$ vs 0.91). Conversely, the component extracted with PCA had a slightly lower correlation ($r=0.88$). These are all in the right direction (positive) and high.

TABLE 1: FACTOR LOADINGS OF 4 G INCREASING ALLELES.

Continents*	Population
SNP	Factor loading
rs9320913_A	0.77
rs11584700_G	0.80
rs4851266_T	0.95
rs236330_C	0.74

Factor scores and average population IQs are reported in table 2.

TABLE 2: FACTOR SCORES OF 4 G INCREASING ALLELES AND PHENOTYPIC IQS.

Continents*	Population	g factor scores	IQ
AFR	Afr.Car.Barbados	-1.2611	83
AFR	US Blacks	-1.2102	85
AFR	Esan Nigeria	-1.4508	71
AFR	Gambian	-1.4472	62
AFR	Luhya Kenya	-1.5391	74
AFR	Mende Sierra Leo	-1.2412	64
AFR	Yoruba	-1.4649	71
HISP	Colombian	-0.1222	83.5
HISP	Mexican LA	0.0216	88
HISP	Peruvian	-0.3041	85
HISP	Puerto Rican	0.0075	83.5
E.ASN	Chinese Dai	1.1828	N/A
E.ASN	HanChineseBejing	1.39	105
E.ASN	HanChineseSouth	1.3038	105
E.ASN	Japanese	1.2297	105
E.ASN	Vietnam	1.5983	99.4
EUR	UtahWhites	0.7559	99
EUR	Finns	0.7143	101
EUR	British	0.8486	100
EUR	Spanish	0.5990	97
EUR	Tuscan Italy	0.5681	99
SAS	Bengali Banglad.	-0.2573	81
SAS	Gujarati Ind. Tx	0.4710	N/A
SAS	Indian Telegu UK	0.0200	N/A
SAS	Punjabi Pakistan	0.1889	84
SAS	Sri Lankan UK	-0.6095	79

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

The correlation between National IQ and factor scores was 0.92 (N=23, p=0.000). Together with the factor loadings, this suggests that this factor represents a signal of polygenic selection on human intelligence and can be used as an indicator of population-level “genotypic intelligence” or “intellectual potential”.

The regression of IQ on the 4 SNPs g factor is plotted in figures 1a and 1b. Inspection of the Q-Q(residuals vs. theoretical quantiles) plot revealed that residuals were normally distributed.

FIGURE 1A. REGRESSION OF NATIONAL IQ ON THE 4 SNPS G FACTOR (LABELS INDICATE POPULATIONS).

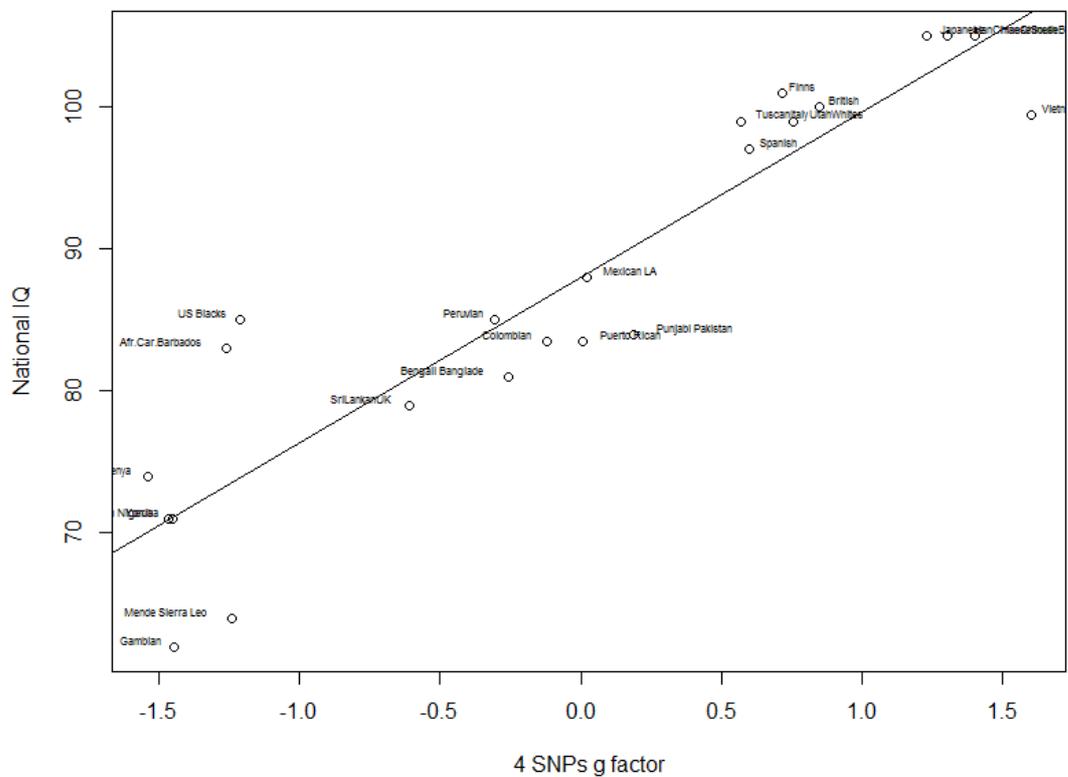
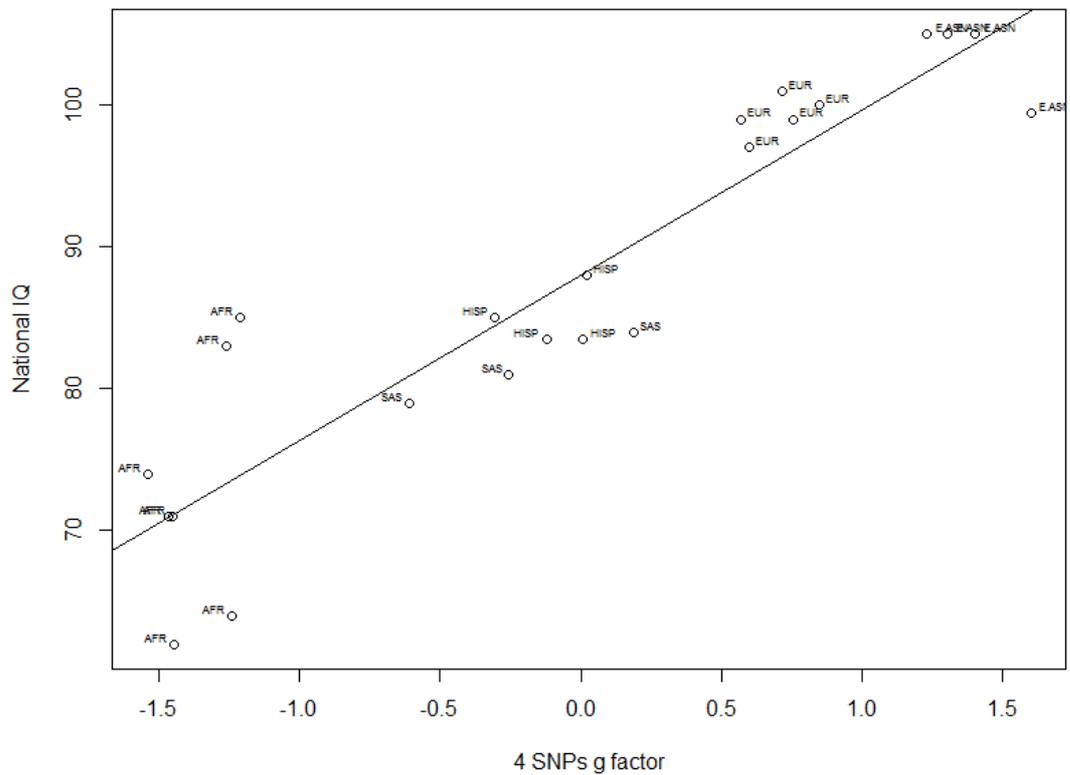


FIGURE 1B. REGRESSION OF NATIONAL IQ ON THE 4 SNPS G FACTOR (LABELS INDICATE CONTINENTAL GROUPS).



Visual inspection of figure 1b shows that populations belonging to the same continent tend to cluster on the genetic factor and that the correlation is driven by racial clusters. South Asians and Hispanics, two groups genetically distant from each other, have similar scores on the genotypic intelligence factor.

A one-way ANOVA was carried out and the difference between racial groups was significant ($F_{4,21}=113.16; p=0.000$)

Results are reported in table 3. Tukey post-hoc test revealed that all the differences between the five groups were significant ($p<0.002$) with the exception of SAS-HISP ($p=0.998$).

TABLE 3: ONE-WAY ANOVA.

	N	Mean	Std. Dev.	Std. Error	95% C.I.	
AFR	7	-1.37	.13	.05	-1.49	-1.25
HISP	4	-0.05	.22	.11	-.39	-.29
EASN	5	1.34	.16	.07	1.14	1.55
EUR	5	.7	.11	.05	.55	.84
SAS	5	0	.43	.19	-.53	.53
Total	26	.01	1.01	.20	-.39	.42

PREDICTING GENOTYPIC IQ

A regression was run with the 4 SNPs g factor as independent and IQs for developed countries only (to eliminate the confounding effect of socioeconomic/environmental disparities) as the dependent variable. This left 9 cases, but the correlation between genetic factor scores and national IQ was stronger compared to the entire sample ($r=0.98$), possibly because the average IQ of developed countries more closely mirrors their genotypic potential. Inspection of the Q-Q(residuals vs. theoretical quantiles) plot revealed that residuals were normally distributed.

To predict genotypic IQs of developing countries (missing from the regression), the unstandardized predicted values were used. A conversion to Greenwich IQ was made by setting the British IQ to 100. These are shown in table 4a.

TABLE 4A. PREDICTED IQS BASED ON REGRESSING IQ OF DEVELOPED COUNTRIES ON THE 4SNPS G FACTOR

	IQ developed countries	Predicted (G.wich) IQ
Afr.Car.Barbados		83.6
US Blacks	85	84.0
Bengali Banglade		91.4
Chinese Dai		102.7
UtahWhites	99	99.3
HanChineseBejing	105	104.3
HanChineseSouth	105	103.6
Colombian		92.5
Esan Nigeria		82.1
Finns	101	99.0
British	100	100.0
Gujarati Ind. Tx		97.1
Gambian		82.1
Spanish	97	98.1
Indian Telegu UK		95.0
Japanese	105	103.0
Vietnam		105.9
Luhya Kenya		81.4
Mende Sierra Leo		83.7
Mexican LA		95.1
Peruvian		91.0

Punjabi Pakistan		94.9
Puerto Rican		93.5
SriLankanUK		88.7
TuscanItaly	99	97.9
Yoruba		82.0

The difference between the predicted (genotypic) and the observed (measured) IQ of developing countries was calculated (table 4b). These are not residuals in the strict sense because the regression analysis was carried out using data for developed countries only. Hence they will be called “pseudoresiduals”. These were correlated to indexes of economic and human development. The correlations were both in the expected direction: $r \times \text{GDP} = -0.34$ ($N=15, p=0.214$); $r \times \text{HDI} = -0.777$ ($N=14, p=0.001$). GDP had an outlier (Puerto Rico), and the correlation increased after its removal ($r = 0.7, N=14, p=0.005$).

TABLE 4B. PREDICTED IQS OF DEVELOPING COUNTRIES, DIFFERENCE PREDICTED-MESURED, PER CAPITA GDP(PPP), HUMAN DEVELOPMENT INDEX (HDI)

	Predicted (G.wich) IQ	“Pseudoresiduals” (Predicted minus measured IQ)	GDP per capita PPP (2010-2013)	HDI (2012)
Afr.Car.Barbados	83.6	0.6	15324	
Bengali Banglade	91.4	10.4	2679	0.554
HanChineseBejing	104.3	-0.7	10485	0.715
HanChineseSouth	103.6	-1.4	10485	0.715
Colombian	92.5	9.0	11540	0.708
Esan Nigeria	82.1	11.1	5303	0.5
Gujarati Ind. Tx	97.1			
Gambian	82.1	20.1	1613	0.438
Indian Telegu UK	95.0			
Vietnam	105.9	6.5	4851	0.635
Luhya Kenya	81.4	7.4	2626	0.531
Mende Sierra Leo	83.7	19.7	1432	0.368
Mexican LA	95.1	7.1	15813	0.755
Peruvian	91.0	6.0	10756	0.734
Punjabi Pakistan	94.9	10.9	4353	0.535
Puerto Rican	93.5	10.0	34183	
SriLankanUK	88.7	9.7		0.745
Yoruba	82.0	11.0	5303	0.5

CONTROLLING FOR THE EFFECT OF MIGRATIONS AND DRIFT

In order to control for the potential confounding effects of migrations and drift on the relationship between IQ and the 4 SNPs g factor, an index of genome-wide genetic distance (Fst) was used. To make the calculations simpler, only the 5 continental groups were used because the Gedmatch distances do not have enough resolution to accurately represent single populations. This is not a major issue as we have seen above that the correlation between national IQs and the 4 SNPs g factor is mostly driven by sub-continental (racial) clusters. As there was not a perfect overlap between Gedmatch and 1000 Genomes clusters (there were more Gedmatch clusters), if a 1000 Genomes group comprised more than one cluster, the average between the sub-clusters was used. This procedure is described in the Appendix. Three separate distance matrixes were created for the dependent (IQ) and the independent (4 SNPs g factor, Gedmatch Distances). These represent the difference (absolute value) between each of the 5 continental groups on the three variables, giving a total of 30 distances (10 for each variable). These are reported in table 5 and the original matrices are reported in table 7 (Appendix).

TABLE 5. DISTANCES BETWEEN 1KG'S FIVE SUB-CONTINENTAL GROUPS

Comparison	4 SNPs g factor difference	Gedmatch distances (Fst)	IQ difference
AFR-HISP	1.32	0.167	12.2
AFR-EASN	2.71	0.164	32.2
AFR-EUR	2.07	0.149	26.2
AFR-SAS	1.37	0.133	8.7
HISP-EASN	1.39	0.116	20
HISP-EUR	0.75	0.045	14
HISP-SAS	0.05	0.087	3.5
EASN-EUR	0.64	0.117	6
EASN-SAS	1.34	0.076	23.5
EUR-SAS	0.7	0.068	17.5

There was a positive correlation between the genome-wide (Gedmatch) distances and the 4 SNPs g factor: $r = 0.67$ (N=10, $p = 0.032$).

The Gedmatch genome-wide distance was not significantly correlated to IQ differences: $r = 0.27$ (N=10, $p = 0.46$). However, the 4 SNPs g factor was significantly correlated to IQ differences: $r = 0.845$ (N=10, $p = 0.002$).

To assess the relationship of the 4 SNPs g factor net of genome-wide distances, a regression was run with IQ difference as dependent and 4 SNPs g factor, Gedmatch distances as independent variables (table 6). A significant model emerged ($F_{2,9} = 26.58$, $p = 0.01$). The 4 SNPs g factor was the only significant predictor (Beta=1.222; $p = 0.005$). Interestingly, the genome-wide distance effect was reversed (compared to the bivariate correlation), implying that greater genome-wide distances are associated with smaller IQ differences between continents.

TABLE 6. REGRESSION OF IQ ON 4 SNPs G FACTOR AND GEDMATCH DISTANCES

Model		Unst.	Stand.	t	Sig.	95.0% C.I. for B		
		Coeff.	Coeff.	Beta		Lower B.	Upper B.	
1	(Constant)	11.841	3.383		2.79	0.027	3.841	19.84
	4 SNPs g factor	14.832	2.122	1.222	6.99	0.005	9.815	19.849
	Gedmatch Dist.	-122.6	38.384	-0.559	-3.92	0.015	-213.43	-31.902

a Dependent Variable: IQ

DISCUSSION

Factor analysis was used to extract a factor from the frequencies of 4 alleles for 26 populations (1000 Genomes). Its interpretation as an indicator of genotypic intelligence or the strength of natural selection on it was supported by a strong correlation ($r=0.92$) to the average phenotypic (national/ethnic) IQs of 23 populations. The four alleles loaded highly and in the expected direction on this factor, supporting its reliability. There were significant sub-continental differences between groups, with a hierarchy topped by East Asians and Europeans, Hispanics and South Asians in the middle and Sub-Saharan Africans at the bottom. Further evidence that the factor represents selection and not genome-wide evolutionary processes, such as random drift or migrations, comes from the finding that the rank of sub-continental genotypic scores of intelligence did not perfectly match measures of genetic distances obtained from neutral markers and was an independent predictor of IQ. The correlation between sub-continental genetic genome-wide distances and the differences in the 4 SNPs g factor was moderately strong ($r=0.67$), suggesting that the 4 SNPs genetic factor contains noise due to genome-wide evolutionary processes (e.g. migrations, drift), not limited to selection for intelligence. However, genetic distances were not significantly correlated to IQ differences ($r= 0.26$) and in the regression model with 4 SNPs g factor, they predicted IQ differences in the opposite direction (Beta= -0.56). That is, after accounting for the effect of the 4 SNPs g factor, greater genome-wide genetic distances were associated with lower IQ differences. However, this effect was not significant. On the other hand, the 4 SNPs g factor emerged as a strong (and significant) positive predictor of IQ differences (Beta= 1.22).

The results also provide preliminary evidence in favor of the hypothesis that poor environmental conditions (i.e. economic and sociocultural) tend to depress national IQ scores. Countries with lower per capita GDP and a lower index of Human Development tended to have larger positive “residuals”, that is the difference between the score predicted by the regression (of IQs for developed countries on the 4 SNPs g factor) and the actually measured IQ was larger in countries with lower GDP and HDI (r around 0.7). Thus, poorer and less developed countries have yet to reach their full intellectual potential.

The results of this study indicate that the gaps in intellectual performance between some populations can be narrowed via adequate improvement of environmental conditions, however the overall pattern of intellectual scores is due to relatively stable and fixed (genetic) factors and cannot be substantially altered.

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. doi:10.1038/mp.2012.184">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>

Gedmatch, 2013. <https://docs.google.com/document/d/1qOBK4xq1K30fXGchPCaJbPLmq5O-PQtlozyZRDka--c/edit?u\nsp=sharing>

Kirkegaard, E. O. W. The international general socioeconomic factor: Factor analyzing international rankings. *Open Differential Psychology*, 2014.

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

Piffer, D. & Lynn, R. 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*, 2014: 246-249. doi: <http://doi.org/10.1016/j.intell.2014.07.006>

Piffer, D. 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*, 2013: 54: 168200.

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. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 2013: 1467-1471. DOI: <http://doi.org/10.1126/science.1235488>

Rindermann, H., Hoang, Q.S.N., Baumeister, A.E.E. Cognitive ability, parenting and instruction in Vietnam and Germany. *Intelligence*, 2013: 366-377. doi: <http://doi.org/10.1016/j.intell.2013.05.011>

United Nations Development Programme (2014). <http://hdr.undp.org/en/content/human-development-index-hdi>

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>

World Bank (2014). GDP per capita, PPP (current international \$)", World Development Indicators database. http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?order=wbapi_data_value_2013+wbap\ni_data_value+wbapi_data_value-last&sort=desc

APPENDIX:

TABLE 7. DISTANCE MATRICES (1KG G FACTOR, GEDMATCH, IQ):

1KG g Factor		AFR	HISP	EASN	EUR	SAS
-1.37	AFR					
-0.05	HISP	-1.32				
1.34	EASN	-2.71	-1.39			
0.7	EUR	-2.07	-0.75	0.64		
0	SAS	-1.37	-0.05	1.34	0.7	

Gedmatch Distances	AFR					
	HISP	0.167				
	EASN	0.164	0.116			
	EUR	0.149	0.045	0.117		
	SAS	0.133	0.087	0.076	0.068	
IQ						
72.8	AFR					
85	HISP	12.2				
105	EASN	-32.2	-20			
99	EUR	-26.2	-14	6		
81.5	SAS	-8.7	3.5	23.5	-17.5	0

Calculation of genome-wide distances (Fst) by sub-continental group using Gedmatch K13 data. Some 1000 Genomes groups comprise a more than one Gedmatch group, and these are reported below. The average of sub-comparisons for each comparison is used in the final calculation.

AFR-HISP

Sub.Sah.Afr-West Med:0.15

Sub.Sah.Afr.-North Atl.:0.146

Sub.Sah.Afr-Amerindian:0.204

HISP-EASN

North Atl.-East Asian:0.114

West Med-East Asian: 0.122

Amerindian-East Asian:0.113

EUR-AFR

North Athl.-S.Sah.Afr:0.146

Baltic-S.Sah.Afr.:0.15

West Med-S.Sah.Afr:0.15

EUR-HISP

West Med-West Med:o

North Atl-North Atl:0

Baltic-Amerindian:0.137

EUR-EASN

NorthAtl-East Asian:0.114

West Med-East Asian:0.122

Baltic-East Asian:0.114

SAS-HISP

North Atl-South Asian:0.064

West Med-South Asian:0.076

Amerindian-South Asian:0.12

EUR-SAS

NorthAtl-South Asian:0.064

Baltic- South Asian: 0.065

West Med-South Asian:0.076