

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

Davide Piffer*

Ulster Institute for Social Research, UK

Weak widespread (polygenic) selection is a mechanism that acts on multiple SNPs simultaneously. The aim of this paper is to suggest a methodology to detect signals of polygenic selection using educational attainment as an example. Educational attainment is a polygenic phenotype, influenced by many genetic variants with small effects. Frequencies of 10 SNPs found to be associated with educational attainment in a recent genome-wide association study were obtained from HapMap, 1000 Genomes and ALFRED. Factor analysis showed that they are strongly statistically associated at the population level, and the resulting factor score was highly related to average population IQ ($r=0.90$). Moreover, allele frequencies were positively correlated with aggregate measures of educational attainment in the population, average IQ, and with two intelligence increasing alleles that had been identified in different studies. This paper provides a simple method for detecting signals of polygenic selection on genes with overlapping phenotypes but located on different chromosomes. The method is therefore different from traditional estimations of linkage disequilibrium.

Key Words: Polygenic; Selection; Educational attainment; Intelligence; SNP; HapMap; 1000 genomes; Race differences.

Introduction

Theory and Hypothesis

Polygenic adaptation (or weak widespread selection) is a model proposed to explain the evolution of highly polygenic traits that are partly determined by common, ancient genetic variation (Pritchard & Di Rienzo, 2010). This type of selection acts on multiple genetic polymorphisms simultaneously. As a

* Address for correspondence: Davide Piffer, Via Molina 15, 54033-Avenza (MS), Italy; email: davidepiffer@libero.it

result, “the effects of polygenic adaptation on patterns of variation are generally modest and spread across many haplotypes across any one locus” (Turchin et al, 2012).

A prediction of polygenic selection is that “the trait-increasing alleles will tend to have greater frequencies in the population with higher trait values, compared to the population with lower trait values” (Turchin et al, 2012). Another prediction of the polygenic selection model (explicitly advanced and tested here for the first time) is that alleles with similar function are statistically associated at the population level, so that populations which have undergone natural selection for a particular trait will have higher frequencies of most alleles associated with that trait, compared to populations upon which selection was weaker, absent or in the opposite direction. Thus, a method of detecting polygenic selection signals is to test statistical associations between allele frequencies of two or more unlinked polymorphic genes (located on different chromosomes) known to be associated with a particular trait within populations, and correlating the allele frequencies with average population trait values (e.g. IQ, height, disease susceptibilities, etc.). As the genes are located on different chromosomes, an explanation in terms of linkage disequilibrium is ruled out.

The present study will apply this method to genetic polymorphisms that are associated with a cognitive phenotype, educational attainment, as an example of a polygenic trait whose genetic variation can be accounted for by a model of weak widespread selection acting on pre-existing genetic variation. An analysis of the distribution of these genetic variants across human populations and their relationships with measures of the phenotype across populations will provide data to test the hypothesis that polygenic selection accounts for their different frequencies

and the observed population differences in educational attainment. As IQ and educational attainment are highly related constructs (Deary et al, 2006; Kaufman et al, 2012), this hypothesis also predicts that the frequencies of many educational attainment alleles correlate with population IQ. Indeed, in a subsample of Rietvald et al's study (2013) for which cognitive test scores were available, the polygenic score of educational alleles explained individual differences in cognitive function, and explained a larger fraction of the variance in cognitive function ($R^2 = 2.5\%$) than in educational attainment for that subsample.

Thus, another prediction is that IQ increasing alleles will be positively correlated with educational attainment increasing alleles. Two single nucleotide polymorphisms (SNPs) whose associations with intelligence seem to be robust because they have been replicated in several independent studies were chosen as representative of intelligence increasing alleles. The first is rs236330, located within gene *FBNPIL*, whose significant association with general intelligence has been 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). The second SNP is rs324650. It was included because its association with IQ has been replicated in four association studies (Comings et al, 2003; Dick et al, 2007; Gosso et al, 2006, 2007). This SNP is located in the gene *CHRM2* (cholinergic receptor, muscarinic #2), which is involved in neuronal excitability, synaptic plasticity and feedback regulation of acetylcholine release.

Educational attainment

Countries around the world differ in their levels of educational attainment, measured either as length of schooling, academic degrees, or performance on scholastic

achievement tests. A variety of factors have been advanced as explanations for these differences. Most commonly these differences are attributed to economic and sociocultural factors. That is, countries with higher human capital invest more in institutional structure and provide higher teaching quality. Conversely, it is not clear whether economic growth leads to higher scores on scholastic achievement tests or vice versa (for a review, see Hanushek and Woessman, 2010).

Importantly, all these explanations fail to take into account genetic variation between individuals and human groups, although most human traits are heritable to a considerable degree (Plomin et al, 2008), and educational attainment is no exception to this general rule. Educational attainment measured as highest degree or length of schooling shows moderate heritability (proportion of variance that is explained by genetic factors) of around 40% (Silventoinen et al, 2004). Moreover, a recent genome-wide association study has identified specific genetic polymorphisms that are responsible for some of the individual variation in educational attainment (Rietveld et al, 2013). Intelligence is a good predictor of performance in educational achievement tests, particularly in subjects such as math and English, where it explained, respectively, 58.6% and 48% of the variance in a longitudinal study based on 70,000+ English children (Deary et al, 2006). Kaufman et al (2012) found high correlations between measures of academic g and cognitive g .

As human groups are different for many morphological traits (Sarich and Miele, 2004) and also for frequencies of functional genetic polymorphisms (e.g. lactose tolerance, hair color, skin pigmentation), it is possible that variation in alleles associated with educational attainment is responsible for many of the observed national, racial or continental differences.

Within the US, there are substantial differences between

racess in the average level of educational attainment. According to the 2003 statistics from the U.S. Census Bureau (Stoops, 2004), Asian Americans had the highest educational attainment, followed by Whites. Blacks and Latinos had the lowest educational attainment. This was particularly evident at the college level. Only 17.3% of Blacks 25 years and older had a Bachelor's degree, compared with 30% of non-Hispanic Whites and 49.8% of Asians. The difference is less pronounced in high school, which is completed by 87.6% of Asians, 89.4% of Whites and 80% of Blacks.

A similar pattern is observed at the country level, particularly on standardized tests of scholastic achievement, such as PISA, where East Asian students (Singapore, China, Japan) consistently obtain higher scores in tests of math, science and reading than their European counterparts. In 2009, China, Korea and Singapore ranked among the top 5. Highest scores were attained by students of Shanghai (China), followed by Korea, Finland, Hong Kong and Singapore.

The high educational attainment of East Asians has often been explained in terms of Confucian values. For example, education professor Yong Zhao stated that it is "no news that the Chinese education system is excellent in preparing outstanding test takers, just like other education systems within the Confucian cultural circle: Singapore, Korea, Japan, and Hong Kong" (Zhao, 2010). This is the typical example of an explanation that goes beyond genetic or even economic factors, instead appealing to historical traditions and cultural values. According to this theory, by placing emphasis on ethics and statecraft, Confucianism fosters high parental interest in education, pressure on children to succeed at school, and the priority it receives in terms of family financial investment, which in turn are responsible for the higher educational attainment of countries dominated by this value system (Starr, 2012).

This paper takes a very different stance, by focusing on the evolutionary and genetic basis of educational attainment.

Methods

Genes associated with educational attainment were obtained from a very recent genome-wide association study relying on a very large sample (126,559 individuals), which identified 10 SNPs associated with educational attainment that reached suggestive genome-wide significance. The outcome measures were an individual's years of schooling and a binary variable for college completion (Rietvald et al., 2013). Three of the 10 SNPs (rs9320913; rs11584700; rs4851266) replicated in a subsequent meta-analysis.

The 10 SNPs were searched on HapMap, release#28 (hapmap.ncbi.nlm.nih.gov) and 1000 Genomes, in order to find allele frequencies for different populations. The frequencies of alleles that had a positive association (beta value) with educational attainment in the combined dataset in Rietvald et al. (2013) are reported in Table 1a and Table 1b. In order to find allele frequencies for more populations, the 10 educational SNPs were searched on ALFRED (The Allele Frequency Database, <http://alfred.med.yale.edu>). Two SNPs (rs13188378 and rs8049439) were found on ALFRED. When an SNP was not found on ALFRED, the most closely linked SNPs was searched, and if not found, the second closest and so on, until $r \geq 0.8$. In this way, a total of 7 SNPs were found, for a total of 50 populations (after list-wise deletion of missing data). Linkage disequilibrium was calculated with SNAP (SNP Annotation and Proxy Search, <https://www.broadinstitute.org/mpg/snap/>), using the 1000 Genomes pilot 1 dataset, CEU as population panel and a distance limit of 500 kB. Table 2 reports the SNPs corresponding to the Rietvald et al. SNPs, and their linkage disequilibrium score (r^2). The frequencies of alleles for the 50 populations are reported in Table 1c.

Data on educational attainment for different races in the US were obtained from the US Census Bureau Report (Stoops, 2004). The PISA (Programme for International Student Assessment) 2009 (www.oecd.org/pisa) scores were used as a measure of country-level educational attainment.

IQs were obtained from Lynn (2006) and Lynn & Vanhanen (2012).

Results

Educational Attainment

Tables 1a and 1b report frequencies of alleles associated with higher educational attainment (“beneficial” for short) for the 14 populations of 1000 Genomes and the 11 HapMap populations (list-wise deletion of missing data). Frequencies of the 10 SNPs were averaged for each population, so as to obtain a single reliable value (or polygenic score) that also avoided the problem of multiple comparisons. The polygenic score is reported in the last column. The results are similar across the HapMap and 1000 Genomes data sets: East Asian populations (Japanese, Chinese) have the highest average frequency of “beneficial” alleles (39%), followed by Europeans (35.5%) and sub-Saharan Africans (16.4%).

Table 3 reports values for measures of educational attainment (PISA score).

Table 4 reports the correlation matrix for the frequencies of the top 3 SNPs (1000 Genomes). HapMap data were excluded as they had many missing values. They are all positive, two are significant ($p < 0.05$, two-tailed t test), and one is nearly significant ($p = 0.051$).

The correlation of the polygenic score (1000 Genomes, Table 1a) with PISA was 0.70 ($p < 0.05$). All correlations between the top three SNPs (rs9320913; rs11584700; rs4851266) and PISA scores were positive ($r = 0.48; 0.87; 0.84$), and the latter two were highly significant ($p < 0.01$).

Since the top 3 SNPs were well correlated among each

other, a principal components analysis (PCA) of frequencies (only 1000 Genomes, as HapMap had many missing data) for all ten SNPs was performed with oblimin rotation, to test the prediction derived from the polygenic selection model, that a positive statistical correlation between frequencies of alleles associated with the same trait should be observed. PCA was chosen on the basis that as molecular measures, the reliabilities of the SNP frequencies are likely to be high. Therefore the error variance amongst variables is low. PCA is sensitive to low variable reliability as it relies on estimating all variance shared between indicators (including error). Oblimin rotation was used to determine whether there is more than one natural factor within the dataset. This is important as the presence of one clear natural factor (as evidenced by a high correlation between rotated factors) might necessitate the use of additional exploratory forms of factor analysis.

Two components were extracted that explained 45.3 and 31.14% of the variance, respectively. Kaiser-Meyer-Olkin was acceptable (0.66) and Cronbach's α was high (0.84). The second factor correlated with the first at 0.05, hence could not be interpreted as part of an overarching general factor. On this basis the oblimin-rotated first principal component was used as a 'natural' factor, hence making allowances for substantial preferential and cross-loading on a second principal component. Finally, this second factor was not clearly interpretable in terms of identity as it had near-zero loadings on two of the top 3 SNPs (rs11585700 and rs4851266). Therefore the analysis is focused on the oblimin-rotated PC1. Factor loadings for each SNP on this PC1 and their standardized factor scores are reported in Table 5a and Table 6, respectively. Factor loadings for four SNPs are very high (0.88-0.97).

Scores of the first factor (PC1) for the 14 populations of

1000 Genomes were correlated with population IQ. The correlation was very high and highly significant ($r=0.90$, $N=14$; $p<.001$). The correlation of PC1 with PISA scores was also high and significant ($r=0.83$; $N= 11$; $p<0.05$).

PC1 also correlated highly with the frequencies of the two IQ increasing alleles (rs236330 C and rs324650 T): $r= 0.838$ and 0.815 , respectively ($p<0.05$).

Moreover, the frequencies of the two IQ increasing alleles were positively correlated with the frequencies of the top 3 educational attainment alleles. For rs236330 C, the correlations with the top three educational attainment alleles were all significant ($r= 0.71$; 0.57 ; 0.86 ; $N=14$; $p<0.05$). Rs324650 T was positively correlated with the three alleles and two correlations were significant ($r= 0.61$; $r=0.82$; $p<0.05$). The polygenic score of the educational attainment alleles was positively correlated with the frequencies of the two intelligence alleles ($r=0.91$; $r= 0.67$; $p<0.05$).

In order to get a more representative sample of world ethnic groups and populations, frequencies taken from ALFRED were used. A factor analysis was carried out with the principal components analysis method and oblimin rotation. Two factors were extracted that explained respectively 45.1 % and 18.9 % of the variance. The two components were uncorrelated ($r= 0.096$), hence could not constitute an overarching factor and the second component was not clearly interpretable. Kaiser-Meyer-Olkin (KMO) was good (0.72). Table 5b reports the structure matrix with the loadings for the first PC. 6 of the 7 SNPs loaded positively and respectably high on the first factor.

Table 1c reports their frequencies for the 50 populations, along with their factor score and estimated racial IQ (Lynn, 2006). As the individual IQs for most tribes/ethnic groups were not available, the factor scores of the ethnic groups were averaged to obtain the factor score for each racial group. The

correlation between racial IQ and racial factor score was $r=0.95$ ($N=9$; $p<0.05$).

Ancestral vs. Derived Alleles

Allele status was checked on dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>). Ancestral alleles are determined by comparison with the chimpanzee genome. Derived alleles are unique to humans, whereas ancestral alleles are shared with chimpanzees. 9 out of the 10 alleles associated with higher educational attainment were derived alleles. The significance of this result was assessed with a binomial calculation, assuming that under a purely neutral mechanism of evolution (no selection), the probability of a derived allele conferring higher educational attainment is 50%. The binomial probability for $X \geq 9$ is 0.01. Thus, this result is highly significant.

Discussion

The national ranking for scores on standardized tests of educational attainment closely mirrors the gradient observed in frequencies of genes associated with this construct. East Asians have the highest frequencies of alleles beneficial to educational attainment (39%) and consistently outperform other racial groups both within the US and around the world, in terms of educational variables such as completion of college degree or results on standardized tests of scholastic achievement. Europeans have slightly lower frequencies of educational attainment alleles (35.5%) and perform slightly worse in terms of educational attainment, compared to East Asians. On the other hand, Africans seem to be disadvantaged both with regards to their level of educational attainment in the US and around the world. Indeed, Africans have the lowest frequencies of alleles associated with educational attainment (16%).

The polygenic score of educational attainment alleles was

a good predictor of PISA scores ($r=0.84$, $p<.05$).

As shown in Table 4, the 3 top SNPs associated with educational attainment were highly correlated among each other. Since the three top SNPs associated with educational attainment are located on different chromosomes (chromosomes 6, 1, and 2, respectively), an explanation of the statistical association in terms of linkage disequilibrium (LD) is not viable.

In fact, factor analysis of 1000 Genomes data confirmed that 6 of the 10 SNPs associated with educational attainment loaded respectably on a single factor (Table 5a). This factor likely represents a nonrandom evolutionary force such as natural selection. Indeed, two of the top 3 SNPs (rs11584700, rs4851266) with the lowest p values in Rietveld et al's meta-analysis (2013) had also the highest factor loadings in this study (0.9 and 0.97) (Table 5a). This implies that the SNPs most strongly associated with educational attainment have also the strongest association with the other SNPs, suggesting that the association is directly proportional to the strength of selection on any given SNP. Table 6 shows factor scores for each population. These can be interpreted as representing a rough estimate of the strength of selection for the phenotype (educational attainment) on each population. Since IQ and educational attainment are highly related constructs, a positive correlation between the examined alleles and population IQ was predicted. The correlation between the first factor and population IQ was very high (0.90), suggesting that it represents a genetic background of higher intelligence. Moreover, frequencies of IQ increasing alleles were highly correlated with frequencies of educational attainment alleles.

The extracted factor reached highest values among East Asians (around 1-1.5), Europeans have a slightly lower factor score (0.1-0.4), and Africans obtained the lowest (negative) factor score (-1.4/-1.6). These results were confirmed and

extended by the analysis of ALFRED's data for 50 populations from all major geographic areas and all continents. The factor scores for 8 racial groups were highly correlated with their estimated IQs ($r=0.95$). Importantly, this analysis also disproves any claims that the factor scores simply represent distance from Africa, as genetically and spatially very distant human groups, such as Native Americans and Oceanians, have lower factor scores than groups that are geographically and genetically closer to Africans, such as the Europeans. Moreover, Native Americans have much lower factor scores than East Asians, despite their high genetic resemblance. This implies that the selective pressures for higher IQ continued after the split between north-east Asian populations and Americans or between South-east Asian populations and Australians.

On the other hand, populations from Central Asia and the Middle East had factor scores comparable to Europeans, suggesting that their lower average IQs can be improved through better environmental conditions (nutrition, schooling, etc.). Finally, the lowest factor scores were observed in the San and Pygmy ethnic groups, which accordingly have the lowest predicted IQ (Lynn, 2006).

Metaphorically, this factor could be seen as a "magnet" attracting all other *unmeasured* educational attainment alleles, located throughout the whole genome. As the effect size of each SNP is typically very low (around 0.1%), even 10 SNPs would not account for more than 1% of the variance in IQ or educational attainment scores across populations. The likely explanation for why the effect size for the 10 SNPs at a cross population level detected in this study is so high (around 80%), is that the alleles are not randomly distributed across human races, so that the combined frequency of a few alleles *predicts* the frequencies of many other alleles affecting the same phenotype. This inflates the correlation with the

phenotype well beyond anything that would be explainable by the modest effect sizes of the examined SNPs.

This is nothing more than the principle applied to psychometric instruments, such as IQ tests or personality scales, where a handful of items produce a reliable score, precisely because these items represent an underlying, latent factor and are thus correlated among each other. Even reliable psychometric scales are usually composed of around 10 items, equal to the number of SNPs examined in the present study, which in turn showed good internal reliability (Cronbach's $\alpha = 0.84$). A model based on random evolution or genetic drift alone cannot account for such a pattern. Indeed, *whenever the phenotypic effects of any set of two or more alleles are similar, population-level correlations suggest co-selection for the same trait*. Let a set of different populations be examined with regards to two polymorphic genes with similar effects on a phenotype. Each gene has two alleles with opposite effects on the same phenotype such that one allele increases the values of the trait (trait increasing) compared to the other allele (trait decreasing). In this case, the capital letter codes for greater value of such a trait: A, a for gene 1 and B, b for gene 2. Under the null hypothesis of random evolution, frequency of allele A is expected to be uncorrelated with the frequencies of alleles B and b. Similarly, frequency of allele a should be uncorrelated with alleles B and b. Thus, under the null hypothesis of random evolution, the expected correlation between trait increasing (A, B) or trait decreasing (a, b) alleles is 0. Instead, if frequency of allele A (trait-increasing) is positively correlated to frequency of allele B (trait-increasing) and negatively correlated to frequency of allele b, then this suggests a mechanism other than neutral evolution (Kimura, 1984) such as natural selection (Piffer, 2013). The support for this inference increases when greater numbers of trait-associated polymorphisms show this pattern. This “positive

manifold” of trait-enhancing alleles can be operationalized by their factor score. Thus, factor scores represent an underlying force of natural selection or a “metagene” for educational attainment, intelligence or related traits spanning SNPs located on different chromosomes.

In fact, a mechanism such as weak widespread selection acting on many alleles likely accounts for these results. Weak widespread selection acts on multiple SNPs simultaneously. As a result, “the effects of polygenic adaptation on patterns of variation are generally modest and spread across many haplotypes across any one locus” (Turchin et al, 2012). Polygenic adaptation (or weak widespread selection) is a model proposed to explain the evolution of highly polygenic traits that are partly determined by common, ancient genetic variation (Pritchard & Di Rienzo, 2010).

Confirming the prediction of the polygenic selection model, this study found that the trait-enhancing alleles had higher frequencies in populations with higher trait values (higher educational attainment and higher IQ). Confirming another prediction of the polygenic selection model, this study found strong statistical association at the population level between alleles known to be associated with the same trait within populations and also other alleles correlated with a similar trait (IQ).

Nine of the 10 alleles associated with educational attainment were derived, thus unique to humans and not shared with non-human primates. This result was significant ($p=0.01$) and is predicted on the basis of the assumption that humans have evolved by natural selection to become more intelligent than their primate cousins. The results show that this evolutionary process, which was already far advanced at the time when modern humans spread across the globe approximately 65,000 years before present, has continued in modern human populations after that time. It invalidates

theories that assume, explicitly or implicitly, that human cognitive evolution has ended with the first appearance of physically modern *Homo sapiens* (e.g., Tooby and Cosmides, 1992).

Conclusion

This paper provides an example of a novel methodology for detecting signals of polygenic selection on more than two SNPs. In the case of a large number of alleles, factor analysis is recommended in order to detect recent selection on polygenic traits.

The method can be used as a tool in gene discovery. Consider any polygenic trait for which some genetic polymorphisms have been discovered already in genome-wide association studies or through genome sequencing, and that shows significant differences between human populations (e.g., height, type 2 diabetes, hypertension, bone mineral density, skin color...). In these cases, a factor score can be calculated from the frequencies in HapMap, 1000 Genomes or ALFRED. This factor score is then correlated with SNP allele frequencies throughout the genome. SNPs that are highly correlated with the factor score can be used in future genome-wide association studies. This strategy can greatly decrease the number of SNPs that are included in a genome-wide association study, reducing the multiple-testing problem and required sample sizes.

Finally, this is the first study to provide systematic (albeit preliminary) evidence that differences between countries and races in IQ and educational attainment are related to genetic factors.

Acknowledgments:

I'd like to thank Michael Woodley and Gerhard Meisenberg for their very helpful discussions and comments.

References

- Benyamin, B., Pourcain, B.St., Davis, O.S., Davies, G., Hansell, N.K., Brion, M.-J.A. et al
(2013). Childhood intelligence is heritable, highly polygenic and associated with *FBNPIL*. *Molecular Psychiatry*, doi:10.1038/mp.2012.184.
- Comings D.E., Wu, S., Rostamkhani, M., McGue, M., Lacono, W.G., Cheng, L.S. & MacMurray, J.P.
(2003). Role of the cholinergic muscarinic 2 receptor (*CHRM2*) gene in cognition. *Molecular Psychiatry* 8: 10-11. doi: 10.1038/sj.mp.4001095
- Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S.E., Liewald, D., Xiayi, K., Le Hellard, S. et al
(2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry* 16: 996-1005.
- Deary, I.J., Strand, S., Smith, P. & Fernandes, C.
(2006). Intelligence and educational attainment. *Intelligence* 35: 13-21.
- Dick, D.M., Aliev, F., Kramer, J., Wang, J.C., Hinrichs, A., Bertelsen, S., Kuperman, S., Schuckit, M., Nurnberger, J. Jr., Edenberg, H.J., Porjesz, B., Begleiter, H., Hesselbrock, V., Goate, A. & Bierut, L.
(2007). Association of *CHRM2* with IQ: converging evidence for a gene influencing intelligence. *Behavior Genetics* 37: 265-272. doi: 10.1007/s10519-006-9131-2.
- Gosso, M.F., van Belzen, M.J., de Geus, E.J.C., Polderman, J.C., Heutink, P., Boomsma, D.I. & Posthuma, D.
(2006). Association between the *CHRM2* gene and intelligence in a sample of 304 Dutch families. *Genes, Brain and Behavior* 5: 577-584.
- Gosso, F.M., de Geus, E.J.C., Polderman, T.J.C., Boomsma, D.I., Posthuma, D. & Heutink, P.
(2007). Exploring the functional role of the *CHRM2* gene in human cognition: results from a dense genotyping and brain expression study. *BMC Medical Genetics* 8: 66.
- Hanushek, E.A. & Woessmann, L.
(2010). *The Economics of International Differences in Educational Achievement*. Discussion paper no. 4925, IZA, Bonn.
- Kaufman, S.B., Reynolds, M.R., Liu, X., Kaufman, A.S. & McGrew, K.S.
(2012). Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence* 40: 123-138.
- Lynn, R.
(2006). *Race Differences in Intelligence: An Evolutionary Analysis*. Augusta GA: Washington Summit.

- Lynn, R. & Vanhanen, T.
(2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.
- Kimura, M.
(1984). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Piffer, D.
(2013). Statistical associations between genetic polymorphisms modulating executive function and intelligence suggest recent selective pressure on cognitive abilities. *Mankind Quarterly*, in press.
- Plomin, R., DeFries, J.C., McClearn, G.E. & McGuffin, P.
(2008). *Behavioral Genetics*, 5th edition. New York: Worth Publishers.
- Pritchard, J.K. & Di Rienzo, A.
(2010). Adaptation—not by sweeps alone. *Nature Reviews Genetics* 11: 665-667.
- Rietveld, C.A., Medland, S.E., Derringer, J., Yang, K., Esko, T. (...) & Koellinger, P.D.
(2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340: 1467-1471.
- Sarich, V. & Miele, F.
(2004). *Race: The Reality of Human Differences*. Westview Press.
- Silventoinen, K., Krueger, R.F., Bouchard, T.J., Kaprio, J. & McGue, M.
(2004). Heritability of body height and educational attainment in an international context: comparison of adult twins in Minnesota and Finland. *American Journal of Human Biology* 16: 544-555.
- Starr, D.
(2012). China and the Confucian Education Model. *Universitas* 21.
- Stoops, N.
(2004). *Educational Attainment in the United States*. U.S. Census Bureau.
- Tooby, J. & Cosmides, L.
(1992). The psychological foundations of culture. In: J.H. Barkow, L. Cosmides & J. Tooby: *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*. New York, Oxford: Oxford University Press.
- Turchin, M.C., Chiang, C.W.K., Palmer, C.D., Sankararaman, S., Reich, D., GIANT consortium & Hirschorn, J.N.
(2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* 44: 1015-1019.
- Zhao, Y.
(2010). A true wake-up call for Arne Duncan: the real reason behind Chinese students' top PISA performance. <http://zhaolearning.com/2010/12/10/a-true-wake-up-call-for-arne-duncan-the-real-reason-behind-chinese-students-top-pisa-performance/>

Table 1a.
Frequency (%) of alleles associated with higher educational attainment (1000 Genomes).*

IQ	rs9320913 (A)	rs3783006 (C)	rs8049439 (T)	rs13188378 (G)	rs11584700 (G)	rs4851266 (T)	rs2054125 (T)	rs3227 (C)	rs4073894 (A)	rs12640626 (A)	Average
AFR	19	35	58	0	8	4	0	12	11	17	16.4
AMR	40	43	58	3	11	27	4	58	12	62	31.8
ASN	39	29	72	1	31	56	0	84	6	73	39.1
EUR	50	42	65	6	23	37	6	48	21	57	35.5
ASW	86	23	50	0	7	9	0	16	7	25	16.9
LWK	74	17	60	0	9	1	0	10	17	17	17.1
YRI	71	18	32	0	6	5	0	11	6	11	15
CLM	83.5	42	53	3	9	23	3	62	22	57	32.9

MXL	88	30	34	56	1	9	30	5	68	5	73	31.1
PUR	83.5	51	43	64	7	15	25	6	42	11	53	31.7
CHB	105.5	42	23	76	1	30	57	0	87	5	75	39.6
CHS	106	40	33	64	1	25	59	0	87	5	75	38.9
JPT	105	35	30	78	0	38	51	0	78	7	70	38.7
CEU	100	49	46	61	8	21	40	6	45	18	56	35
FIN	97	52	33	61	6	27	35	10	59	25	57	36.5
GBR	100	49	44	67	6	26	41	8	40	18	61	36
IBS	97	43	54	71	0	21	29	4	71	21	43	35.7
TSI	100	52	43	69	5	18	34	3	45	23	59	35.1

* AFR: African; AMR: American; ASN: Asian; EUR: European; ASW: African ancestry in SW USA; LWK: Luhya, Kenya; YRI: Yoruba, Nigeria; CLM: Colombian; MXL: Mexican ancestry from LA, California; PUR: Puerto Ricans from Puerto Rico; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese; JPT: Japanese in Tokyo, Japan; CEU: Utah Residents with Northern and Western European Ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian population in Spain; TSI: Toscani in Italy.

Table 1b.
Alleles associated with higher educational attainment (HapMap).*

	rs9320913 (A)	rs3783006 (C)	rs8049439 (T)	rs13188378 (G)	rs11584700 (G)	rs4851266 (T)	rs2054125 (T)	rs3227 (C)	rs4073894 (A)	rs12640626 (A)	Average
ASW		0.482		0.07	0.079		0.088		0.219		
CEU	0.508	0.375	0.615	0.069	0.208	0.414	0.058	0.415	0.164	0.58	0.341
CHB	0.419	0.167	0.73	0.015	0.35	0.559	0	0.859	0.073	0.748	0.392
CHD		0.739		0.321	0.556			0.055		0.752	
GHI		0.718		0.272	0.33			0.119		0.569	
JPT	0.326	0.311	0.783	0.004	0.358	0.509	0	0.85	0.08	0.735	0.396
LWK		0.595		0.086	0.023			0.182		0.168	
MEX		0.595		0.078	0.336			0.043		0.741	

MKK	0.596	0.032	0.093	0.228	0.266
TSI	0.701	0.176	0.368	0.225	0.578
YRI	0.189	0.333	0.571	0	0.148

* ASW: African ancestry in Southwest USA, CEU: Utah residents with Northern and Western European ancestry from the CEPH collection, CHB: Han Chinese in Beijing, China, CHD: Chinese in Metropolitan Denver, Colorado, GIH: Gujarati Indians in Houston, Texas, JPT: Japanese in Tokyo, Japan, LWK: Luhya in Webuye, Kenya, MEX: Mexican ancestry in Los Angeles, California, MKK: Maasai in Kinyawa, Kenya, TSI: Tuscan in Italy, YRI: Yoruban in Ibadan, Nigeria.

Table 1c.
Alleles associated with higher educational attainment (ALFRED).*

	rs1906252 T	rs8049439 T	rs13188378 G	rs11588857 A	rs11686372 A	rs2966 T	rs4073643 A	PC1	IQ
Africa									
<i>Pygmy and Bushmen</i>									
Bantu	8.5	42.5	0	0	0	17.5	8.5	-1.89	
San	8	33	0	0	0	0	0	-2.29	
Biaka	21	23	0	0	10	18	2	-1.89	
Mbuti	13	13	0	3	3	0	3	-2.24	
<i>Western Africa</i>									
Yoruba	23	58	0	2	0	10	10	-1.65	71
Mandenka	27	48	0	0	13	2	10	-1.60	

Bantu	8.5	42.5	0	0	0	17.5	8.5	-1.89
Middle East								<i>-0.19</i> 92 ^a
Mozabite	57	75	2	10	30	33	22	-0.24
Bedouin	38	86	0	15	29	43	18	-0.19
Druze	51	65	1	18.5	36	60	31.5	0.33
Palestinian	58	77	1	16	28	35	26	0
Europe								0 100
Adygei	38	74	0	26.5	41	53	26.5	0.36
Basque	40	67	8	15	31	54	23	-0.30
French	48	69	3	19	38	40	24	-0.01
Italians	46	72	2	25	40.5	37.5	24	0.14
Orcadian	56	44	13	31	47	59	16	-0.03

Russian	32	56	2	14	32	68	30	-0.03
Sardinian	25	59	0	14	36	34	13	-0.59
Central Asia								<i>0.1</i> 97 ^b
Burusho	44	78	0	24	28	74	14	0.16
Kalash	38	72	4	34	29	54	22	0.11
Pashtun	33	74	0	13	20	65	22	-0.21
Balochi	24	70	1	25	42	68	22	0.22
Brahui	42	72	0	14	28	76	20	0.06
Hazara	60	65	0	23	50	60	15	0.41
Sindhi	32	84	0	10	26	64	16	-0.23
East Asia								<i>0.97</i> 105
Dai	45	65	0	20	65	85	25	0.87

Mongolia	20	55	0	45	65	70	45	1.25
Daur	61	61	0	50	50	78	39	1.49
Han	41	70	0	31	56	89	26	0.98
Hezhe	28	67	0	56	50	83	11	0.85
Japanese	36	80	0	45	45	84	35.5	1.23
Koreans	43.5	72	0	39.8	50.9	85.2	40.7	1.34
Lahu	40	70	0	15	65	100	15	0.72
Miao	60	60	0	20	65	90	25	1.02
Naxi	33	72	0	28	28	100	22	0.48
Oroquen	35	55	0	20	55	85	40	0.84
She	35	75	0	15	65	90	25	0.81
Tu	45	65	0	30	50	80	35	0.96

Pima, Mexico	38	18	0	0	34	80	0	-0.88
Maya, Yucatan	8	40	2	2	26	84	6	-0.98
Amerindians	27	0	0	0	12	96	19	-0.93
Karitiana	2	2	0	0	50	71	0	-1.17
Surui	24	5	0	0	26	88	0	-1.15

* Allele frequencies, factor scores and estimated IQ. IQs are reported for continental groups, with the exception of Africa, where there is higher genetic variation between ethnic groups. Factor scores for continents/races are average of the populations belonging to each category.

a) 92 is the IQ for Turkish and Moroccan people living in Europe, which is higher than the IQ of the indigenous populations (Lynn, 2006, p.86).

b) Indian and Pakistani children resident in Britain for four or more years. This is higher than the estimated IQ for these ethnicities living in their home countries (Lynn, 2006, pp.82-84).

c) 93 is the IQ of Southeast Asians in the U.S. (Lynn, 2006, p.100), six points higher than that of indigenous Southeast Asians (87).

Table 2.
Educational Attainment SNPs found on ALFRED.

Rietveld et al. (2013)	SNPs in LD ($r^2 \geq 0.8$)
rs9320913	rs1906252 ($r^2 = 0.905$)
rs8049439 (on ALFRED)	
rs13188378 (on ALFRED)	
rs3783006	None found
rs11584700	rs1158857 ($r^2 = 0.866$)
rs4851266	rs11686372 ($r^2 = 1$)
rs2054125	None found
rs3227	rs2966 ($r^2 = 1$)
rs4073894	rs4073643 ($r^2 = 0.901$)
rs12640626	None found

Table 3. PISA scores, average of math, science and reading.

Populations*	PISA 2009
African American (ASW)	422
White American (CEU)	520
Chinese Shanghai (CHSh)	577
Chinese Hong Kong (CHH)	545
Japan (JPT)	539
Mexico (MXL)	420
Great Britain (GBR)	500
Finland (FIN)	543
Spain (IBS)	484
Italy (TSI)	493
Colombia (CLM)	398

* CHS (Hong Kong); CHB; Shanghai. Sub-populations of the US were averages of Math 2003, Science 2006 and Reading 2009, as racial scores were not revealed for all groups in the 2009 PISA reports.

Table 4.
 Correlation matrix for the top 3 educational attainment SNPs. N= 14 populations from the 1000
 Genomes project.

	rs9320913	rs11584700	rs4851266
rs9320913	1	0.533	0.609
rs1158470		1	0.849
rs4851266			1

Table 5a. Structure matrix (factor loadings on oblimin-rotated first principal component). 1000 Genomes data. N= 14 populations.

SNPs	PC1
rs9320913(A)	0.62
rs3783006(C)	-0.26
rs8049439 (T)	0.74
rs13188378 (G)	0.17
rs11584700(G)	0.90
rs4851266 (T)	0.97
rs2054125 (T)	0.14
rs3227 (C)	0.88
rs4073894 (A)	-0.11
rs12640626 (A)	0.89

Table 5b. Structure matrix (factor loadings on oblimin-rotated first principal component), ALFRED data. N= 50 populations.

SNPs	PC1
Rs1906252 T	0.570
Rs8049439 T	0.629
Rs13188378 G	-0.037
Rs11588857 A	.801
Rs11686372 A	.850
Rs2966 T	.664
Rs4073643 A	.761

Table 6. Factor scores of the 1000 Genomes populations. IQ is from Lynn, 2006 and Lynn & Vanhanen, 2012.

Population	IQ	PC1 Scores
ASW	86	-1.439
LWK	74	-1.597
YRI	71	-1.483
CLM	83.5	-0.526
MXL	88	-0.055
PUR	83.5	-0.095
CHB	105.5	1.539
CHS	106	1.084
JPT	105	1.407
CEU	100	0.114
FIN	97	0.394
GBR	100	0.386
IBS	97	0.084
TSI	100	0.186