Genomic ancestry, cognitive ability and socioeconomic outcomes

Emil O. W. Kirkegaard¹

Gerhard Meisenberg²

John Fuerst³

Abstract

Self-identified race/ethnicity (SIRE) groups are known to differ in average cognitive ability, socioeconomic status and genomic ancestry, but the causal paths are disputed. We used data from the Pediatric Imaging, Neurocognition, and Genetics Study (PING; n = 1,369) to examine the relationship between cognitive ability, socioeconomic outcomes and genomic ancestry in the United States.

In regression models using 4 different codings of SIRE as a covariate, we found non-trivial relationships between genomic ancestry and outcomes. The relationship between genomic ancestry and cognitive ability was reduced but was still substantial when parental socioeconomic status was controlled for. Interpretation of this is unclear due to possible genetic confounding. Relationships also held when Hispanics and African Americans were analyzed alone.

In this sample, the mean IQ by ancestry was approximately 100 for European, 78 for African, and 86 for Amerindian. The sample sizes for the remaining ancestries were too small to produce reliable estimates. The results are consistent with genetic models of cognitive and social inequality, but also with certain environmental models. More sophisticated studies are possible and necessary to resolve the question.

Key words: race, ethnicity, biogeographic ancestry, ancestry, admixture, genomics, race differences, group differences, intelligence, IQ, cognitive ability, social status, inequality

1 Ulster Institute for Social Research. Email: emil@emilkirkegaard.dk

2 Professor of Physiology and Biochemistry, Ross University, Dominica. Email: GMeisenberg@rossu.edu

3 Ulster Institute for Social Research. Email: j122177@hotmail.com

1. Introduction

Cognitive ability⁴ and social status are well known to differ by racial groups across the globe (Fuerst & Kirkegaard, 2016b, 2016a; Lynn, 2015; P. L. Roth, Bevier, Bobko, Switzer, & Tyler, 2001). The differences are fairly consistent across countries and time periods. Mean IQ for a number of racial groups are shown in Table 1.

Racial group	Mean IQ
Northeast Asia	105
Europe	100
Central Asia	97
Southeast Asia	93
Middle East	92
Native American	86
Oceania	82.5
Sub-Saharan Africa	71
Pygmy and Bushmen	54

Table 1: Mean cognitive ability by racial group. After Lynn (2015).

The results in Table 1 are based on a mixture of data from traditional IQ tests, Piagetian tests, educational/scholastic tests (PISA, TIMSS etc.). In general, it does not matter much which exact data source is used (Lynn & Meisenberg, 2010; Rindermann, 2007; Sandefur, 2016).

IQ tests were invented in the early 1900s, but there are crude measures of cognitive ability going back a few hundred years. These are based on two main sources of data: writings and numbers. For legal writings, it can be determined whether someone was able to sign it themselves, or had a notary do it for them. This thus functions as a crude measure of (very minimal) literacy (Clark, 2007). For numbers, estimates are based on the observed amount of age heaping, which is the tendency of innumerate people to round their ages to the nearest 5 or 10 because they are unable to calculate the correct age using their year of birth and the current year (A'Hearn, Baten, & Crayen, 2009). This is a very easy test of mathematical ability (subtraction). Thus, one collects historical documents of self-reported age, especially from censuses and tomb stones, and estimates the amount of age heaping. Figure 1 shows the relationship between Lynn's (2012) IQ estimates and age heaping scores in the last part of the 1800s.

4 We use the term *cognitive ability* instead of various other alternatives (intelligence, general intelligence, g, general cognitive ability, general mental ability etc.) in line with many others. For the purpose of this study, the finer distinctions between them are not important. The reader is invited to mentally substitute his own preferred term.

Figure 1: Scatterplots of national IQs (Lynn & Vanhanen, 2012) and age heaping scores (Baten, 2015). As can be there, there are substantial correlations between the old, crude estimates and the modern, much more sophisticated estimates. The relationship becomes weaker over time as an artifact of the increasing ceiling effect (~100% of population has reached the basic competence required per the FLynn effect).

1.1. Evolutionary models of selection for cognitive ability

The topic of evolutionary models for polygenic selection is very complex and spans multiple academic fields including anthropology, archaeology, evolutionary genetics, population genetics, behavioral genetics and genomics. It is not possible to review in depth all the relevant findings here and as such only a brief review is given.

Human population genetic history is incredibly complex and comprises a large number of divergence events and processes. Present day human populations are distinguishable using genomic data, but are not distinct enough to be different species (see Woodley, 2010 for discussion). The exact terminology used for them (race, subspecies, variety, population, genetic cluster, (biogeographic) ancestry group, lineage population etc.) is a matter of on-going debate, but need not concern us here (Andreasen, 2000; Bruer, 1990; Fuerst, 2017; Kaplan & Winther, 2013; Kitcher, 2007; Sesardic, 2010). No matter the terminology chosen, human population genetic differences are real, and since human groups have been relatively or completely isolated in varied environments for many thousands of years, it is not just

possibly, but likely for evolutionary forces to have caused genetic divergence between them (Cochran & Harpending, 2010).

1.1.1. Genomic studies selection in humans

There are several fairly well documented cases of fairly simple recent selection in humans. The first involves the ability of adults to digest milk. This ability usually stops in early childhood, but in some human groups it persists into adulthood for most members. This is congruent with the same groups' use of dairy farming, and thus represents a case of culture-evolution co-evolution (Cochran & Harpending, 2010, p. 77ff; Gerbault et al., 2011).

The second involves the evolution of skin tone⁵ – both dark and light (Jablonski & Chaplin, 2010). Since our closest cousin species are light-skinned under the fur/hair, dark skin must have evolved in humans after the split from our common ancestor and in response to the loss of hair. Later on, when some humans wandered out of Africa and into less sun-exposed areas, they evolved light skin again, though the exact reasons for this are surprisingly murky – the traditionally offered explanations in terms vitamin D and skin cancer protection do not withstand scrutiny (Elias & Williams, 2015). Interestingly, the genetic variants responsible for the light skin of East Asians is not the same as that responsible for Northern Europeans. This illustrates the ability of evolution to solve the same problem in multiple genetically different ways (convergent evolution).

The third case involves high-altitude populations (Beall, 2014). Living in high altitude causes extra stress due to among other things, the reduced oxygen pressure of the air.⁶ Three groups are known to have evolved adaptations to high-altitude living: Tibetans of the Himalaya (Asia, Tibet region of China as well as neighboring countries), Andeans of the Andes mountains (South America, overlapping the territory of Bolivia, Peru and Chile), and Ethiopian highlanders (Northeast Africa, Ethiopia). When the genetics were examined, it turned out that they had evolved different adaptations, though they are all related to oxygen transfer in the body.

The study of highly polygenic selection via changes in frequencies of many alleles (soft sweeps) in humans is quite recent, owing to the recent availability of large-scale powerful computer clusters and large public or semi-public genomic datasets made possible by the falling price of genotyping. Despite this, soft sweeps are considered by some researchers to be the dominant mode of adaptation in humans (Schridder & Kern, 2017). A number of recent studies based on genomic data have found evidence of recent polygenic selection in humans (Berg, Zhang, & Coop, 2017; Joshi et al., 2015; Mathieson et al., 2015; Woodley, Younuskunju, Balan, & Piffer, 2017), including selection that differed by population such that populations would, over time, acquire different average polygenic scores for the traits (Berg et al., 2017; Joshi et al., 2015; Piffer, 2015; Robinson et al., 2015; Turchin et al., 2012). One example of this is the study by Robinsen et al (2015), which examined differences in height and BMI among 14 European national populations and found that some of the observed national differences in height and

5 A variety of terms are used for this trait including *skin color*, *skin tone*, *skin brightness*, *skin lightness*.

6 The oxygen content (volume %) is about constant at human-livable altitudes, however because the air pressure is lower, fewer oxygen molecules are present in a given volume, less oxygen is absorbed with each breath.

BMI could be attributed to genetic causes. Furthermore, it has been shown that the Neanderthal admixture % in modern Eurasians has been decreasing over time, consistent with polygenic selection against whatever traits Neanderthal ancestry has higher polygenic scores for (Gregory et al., 2017; Simonti et al., 2016). Especially notable among the studies mentioned is that some of them found evidence of recent selection for educational attainment (Beauchamp, 2016; Kong et al., 2017; Woodley et al., 2017), which is highly genetically correlated with cognitive ability and often used as a proxy in genomic studies (Hill, Davies, McIntosh, Gale, & Deary, 2017; Marioni et al., 2014; Rietveld et al., 2014; Zabaneh et al., 2017). Taken together, these facts suggest the possibility that selection for cognitive ability has not been equal for all groups in recent history (e.g. last 10k years). There is no *a priori* reason to think that the strength of selection was exactly equal for cognitive ability in the diverse environments humans have found themselves in in the last many thousand years.

1.1.2. Proposed theoretical models

Before reviewing the theoretical models for divergence of human populations in cognitive ability, it is worth reviewing the pre-out of Africa models for human cognitive ability. The pre-out of Africa human brain size has increased dramatically since the split from our common ancestor with the Chimpanzee, as shown in Figure 2.

Figure 2: Cranial capacity in fossil hominids over time. Extant chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens sapiens*) are included for comparison. From Schoenemann (2006).

The increase in brain size is generally considered to reflect an increase in cognitive ability and related cultural sophistication. This is supported by a meta-analysis modern MRI studies, which show that cognitive ability and brain size correlates at about .40 (Gignac & Bates, 2017, corrected for range restriction, data from studies using highly reliable IQ measurements). But how did humans get so smart compared to our Chimpanzee-like ancestors? A variety of proposals have been given (see e.g. reviews in Byrne, 1995; Geary, 2004), and we only briefly review some of the major models.

The social complexity/Machiavellian hypothesis (Dunbar, 2003) posits that navigating human social networks becomes increasingly difficult as they grow in complexity and size. This rewards persons who are better able to do so e.g. by forming crucial alliances to avoid getting killed, or by effective resource collection through diversification of labor. This then leads to increasingly smarter generations that result in increasingly complex social networks. This evolutionary positive feedback loop thus results in a run-away selection for increased intelligence. Support for this comes from a variety of studies showing that social complexity is related to larger brains and higher intelligence, both in studies of non-human animals, primates, and human populations. Furthermore, intelligence is known to be related to stereotype accuracy (Jussim, 2012; Kirkegaard & Bjerrekær, 2016; Lick, Alter, & Freeman,

2017), which leads to more accurate views and hence predictions of other persons' behavior, thus facilitating optimal social behavior towards them. An example of such relevant predictions is likelihood of a person reciprocating a favor or cooperating in way, behavior that is also known to be linked to cognitive ability (evidence reviewed in Jones, 2016; Rindermann, 2018).

The tool use/technology model (Reader & Laland, 2002) posits that increased cognitive ability allowed for innovations in and more adept use of tools and which helped in resource extraction (e.g. extracting termites using a branch). A study of 116 of the 203 known primate species shows that tool use is associated with greater brain size (relative and absolute, controlling for phylogeny). A related model is that of behavioral flexibility which posits that increased cognitive ability enables greater flexibility in behavior which then allows for more efficient resource extraction (e.g. eating a new type of food). Support for this comes from both studies of birds and primates (Overington, Morand-Ferron, Boogert, & Lefebvre, 2009).

The social learning/cultural intelligence hypothesis (Schaik & Burkart, 2011) posits that persons with greater ability to learn from others have improved outcomes because they are able to forage more effectively. One example is tool-assisted nut cracking in Chimpanzees which is learned behavior from others, not rediscovered by each individual. This relies on tool use itself being useful, so it works together with the general tool use hypothesis. Here it can also be noted that in humans, cognitive tests that are more heavily loaded on cultural knowledge show higher heritabilities (Kan, Wicherts, Dolan, & van der Maas, 2013), a finding that seems vaguely congruent with the cultural intelligence hypothesis.

There is some evidence that all of these factors come together because when analyzed across extant primate species, there are positive intercorrelations between tool use, behavioral flexibility, social complexity, social learning and asocial learning, giving rise to a kind of aggregate-level G factor (Fernandes, Woodley, & te Nijenhuis, 2014).

1.1.2.1 Post-out of Africa human divergence in cognitive ability

While the topic is controversial, a number of researchers have proposed theoretical models of recent human evolutionary divergence in cognitive ability. Perhaps the most widely proposed model is cold winter's theory, which is simply the idea that surviving in colder climates is more cognitively demanding than living in warmer climates and hence there should be (additional) selection for cognitive ability (Lynn, 2015; Rushton, 2000; Templer & Arikawa, 2006; Templer & Stephens, 2014). The model came out by comparing national climate data with national IQs which revealed that winter temperatures were the strongest correlate. However, given the limited number of countries such differences in correlations will not generally be statistically reliable, and hence only provide weak to moderate strong support for a cold winter theory model over alternatives such as yearly temperature variation (Ash & Gallup, 2012; Templer & Stephens, 2014).

Another issue with the analysis of national IQs and its environmental correlates is that human populations have not been staying in one place for the last thousands of years, rather they have moved around and admixed with other populations in complicated ways (Reich, 2018). This raises the

possibility that the homelands of current national populations – the usual units of these analyses – are significantly different from their environment of evolutionary adaptedness (EEA). One potential mitigating factor would be if the speed of evolutionary adaption for cognitive ability was quite rapid, in which case it is unlikely that a national population group has ended up in an environment significantly different from their EEA. One suggestive piece of evidence that populations have moved around significantly in this respect is that skin tone has a stronger correlation at the national-level with IQ (about .90) than many of the climatic variables (.50 to .70), suggesting that skin tone tracks recent the relevant recent EEA better than current climatic data. This is based on the assumption that sun exposure is strongly related to the relevant climate, since skin tone is a function of sun exposure in some fashion (Elias & Williams, 2015; Jablonski & Chaplin, 2010).

Another factor not often mentioned but potentially significant is biomes, as seen in Figure 3. An area with the same, say, mean winter temperature might differ noticeably in biome prevalence, which is a likely factor in evolution because it is strongly related to the behavior necessary to survive in the habitat. Unfortunately, given the small sample size and the multitude of plausible evolutionary environment candidates, it seems very difficult to determine which are the relevant ones, if any. One possible solution is to use subnational datasets which offer much larger sample sizes (Kirkegaard, 2016b; Lynn & Yadav, 2015; Manrique Millones, Flores-Mendoza, & Millones Rivalles, 2015), but also likely more strongly violate the assumption of no recent population migration.

Figure 3: Global biome map (Natural Resources Conservation Service, 2003).

Though not often mentioned, a large body of research has investigated cognitive ability – usually proxied by brain size measures – and environmental/climate correlates in non-human animals. The most useful species to study for this are those that enjoy a near-global habitat, are easy to study and has a larger number of species. Given these constraints, birds have been a primary choice of study (Sayol et al., 2016; Sol et al., 2010; Sol, Lefebvre, & Rodríguez-Teijeiro, 2005; Sol, Székely, Liker, & Lefebvre, 2007). It is important to note that one must be careful to not mix migratory and stationary birds in the studies indiscriminately (Sol et al., 2010). This is because migratory birds are not subjected to harsh winters, but are instead subjected to selection for smaller body mass since this lowers the energy cost when migrating (long flights). While ancestral (and a few contemporary) human groups do migrate, they do not migrate as far as migratory birds due to the quite limited ability of humans to travel so far effectively before the advent of modern technology (populations that relied on boats might be an exception). As such, humans have more in common with the non-migratory birds who have to find a way to deal with the colder winter environment. Studies of birds have revealed the expected patterns, namely that birds that live further north and in more seasonally affected areas have larger brains (controlling for body size), more flexible behavior, and more innovative behavior (T. C. Roth & Pravosudov, 2009; Sayol et al., 2016; Sol et al., 2005).

Returning to humans, owing to a general dearth of measures of cognitive ability before about 1900, brain size and its primary proxy cranial capacity has generally been relied upon. Earlier we saw how humans massively increased their brain size during the last 3 million years of evolution. Do measures of brain size also support differential human cognitive ability evolution since out of Africa? Yes, several studies of fossils show that the cranium capacity and pelvis width (needed for birth of larger brains) is related to latitude (Beals et al., 1984). Furthermore, among hunter-gatherer societies, tool complexity (number of different tools) increases with (absolute) latitude (Collared, Kemery, & Banks, 2005).

To summarize, the general thrust of findings in both animal and human ecology is that populations differ in brain size (controlling for body size), cognitive ability measured or proxied in various ways such as by behavioral complexity, social complexity, innovative behavior, tool use. Furthermore, these differences are generally found to be related to plausible environmental/climatic factors such as winter temperature, winter residency (birds), as well as latitude taken as a broad proxy. The fact that these environmental factors are widely regarded as plausible causes for differences in cognitive ability and brain size in non-humans, but strongly contested in humans suggests non-scientific reasons for the dispute (Cofnas, 2015).

1.1.3. Non-evolutionary models for large-scale variation in human cognitive ability

While the focus of this article is on evolutionary models for racial or population differences in cognitive ability, it is important to note that many non-evolutionary models have also been proposed and survey research indicate they are enjoy widespread support among experts (Rindermann, Becker, & Coyle, 2016). Notable proposed causes includes education (equality and quantity), wealth, health/nutrition (Lynn, 2015), aspects of geography and climate e.g. sun exposure; (León & Burga-León, 2018), parasite infections (Eppig, Fincher, & Thornhill, 2010), discrimination, legacy of slavery, culture (Sowell, 1995). Common to all of these proposals is that they have not generally be compared directly with genetic ones, and many of them have a strong *ad hoc* nature that has not dissipated over the years (Jensen, 1973).

1.2. Genomic ancestry and self-identification

When diverse groups of people have their genomes measured (genotyped) and the resulting data analyzed, it is clear that there is structure in the data (Cavalli-Sforza, Menozzi, & Piazza, 1994; Li et al., 2008; Reich, 2018). It has become commonplace in medical genetics and cosumer genetics (23andme, ancestry.com etc.) to genotype persons in order to estimate their ancestry. For clarity, we will refer to this as *genomic ancestry*.⁷

Usually, not all racial groups are present within a given society. In the USA, people of European, African, Amerindian, and Northeast Asian ancestry are present in significant numbers, as well as

⁷ A variety of terms exist in the literature. The most common in the medical genetics literature is *biogeographical ancestry*, but we find it to be long and uninformative. Other alternatives include bioancestry, racial ancestry, and genetic ancestry.

various mixed persons. The presence of mixed persons complicates analysis because one person's label may differ from the next's despite identical ancestry proportions. Furthermore, social scientists, pollsters etc. do not generally collect genetic data from people. Instead, people are asked to classify themselves into various groups and in different ways. We refer to this as *self-identified race/ethnicity* (SIRE).

The relationship between genomic ancestry and SIRE varies by both SIRE and by time and place, even within countries. For instance, in the USA, Whites have been found to have nearly purely European ancestry (96.3%, and trace amounts of other groups), while African Americans have approximately 20% European ancestry and 80% African (Shriver et al., 2003). In contrast, Whites (Branco) in Brazil were on average about 80% European, while Blacks (Pretos) were about 20%. Furthermore, Whites in the south of Brazil were more European (87%) than Whites in the northeast of Brazil (65%) (Fuerst & Kirkegaard, 2016b, supplementary file 13; see also Bryc, Durand, Macpherson, Reich, & Mountain, 2015). This inconstancy of ancestry means that one cannot assume that a given SIRE has the same social meaning in different countries or regions within countries.

1.3. Genomic ancestry and causality

Medical geneticists routinely collect genomic data from study participants because they believe it may be relevant to the observed disparities in various diseases seen between racial groups. In general, they treat relationships between genomic ancestry and disease status as evidence of or suggesting genetic causation, especially when confounders are controlled in regression analysis. Consider some examples:

“The role of genetic predisposition in this disparity is supported by two admixture mapping studies of AAs which demonstrated that greater proportion of European ancestry was inversely associated with fibroids in AA women.” (Giri et al., 2017)

“African Americans in the United States are variably admixed with African and European ancestry (Parra et al., 2001). Due to this heterogeneity, genetic admixture analysis offers a unique opportunity for studying the role of genetic factors within a single, admixed population, controlling for socioeconomic and behavioral factors as well as comorbidities (Pritchard et al., 2000).” (Bidulescu et al., 2014)

“A greater proportion of African genetic ancestry is independently associated with higher FG levels in a non-diabetic community-based cohort, even accounting for other ancestry proportions, obesity and SES. The results suggest that differences between African-Americans and whites in type 2 diabetes risk may include genetically mediated differences in glucose homeostasis.” (Meigs et al., 2014)

Probably hundreds of other such comments can be found in the medical genetics literature. At the time of writing (2017, August), A Google Scholar search for “‘biogeographical ancestry’ disease’ returned 807 hits.⁸

8 <https://scholar.google.dk/scholar?hl=en&q=%22biogeographical+ancestry%22+disease>

Relating genomic ancestry to outcomes is called *admixture analysis* (or association analysis), and given its prominent role in this study and the likely unfamiliarity of the readers with it, we will explain the theoretical basis for it in some detail.⁹ Doing so requires going through a series of simulations that incrementally increase in complexity and realism. Readers who prefer more detailed outputs or are curious about the code basis can find the entire R markdown output in the supplementary materials.

1.3.1. Simulation 1

Initially, we begin with a simple scenario. Suppose there is a trait which is 100% heritable (genetic variation explains 100% of the phenotype variation) and is controlled equally by 10 variants (places of variation in the genome) which can be either A_0 or A_1 , with effect sizes of 0 and 1. Suppose further that each variant has a frequency of 50%, which is unrelated to the frequency of the other variants (i.e. complete linkage equilibrium; no correlations between variant frequencies between populations). Additionally, we assume complete additivity, i.e. that variants can simply be added up; that there are no interactions. Furthermore, we assume there is just a single long genome with one copy for each locus (haploidity). Finally, suppose we have two populations both with 50,000 people. If we simulate data based on these assumptions¹⁰, calculate each individual's polygenic score (sum of the variants they have), then the resulting distributions will look like those in Figure 4.

9 One critic went so far as to claim that “For all the hereditarians’ idle intuitions about differences being part genetic and part environmental, where is the empirical or quantitative theory that describes how this apportioning is supposed to work? There is no such thing as a “group heritability coefficient,” no way to put any meat on the speculative bones about partial genetic determination. In the absence of an actual empirical theory, the discussion is all about your intuitions against my intuitions.” (Turkheimer, 2017).

10 We are also assuming a haploid population, i.e. single copies of each locus. This assumption has no effect on the results because we also assume complete additivity i.e. the effects of different variants can be added up.

Figure 4: Simulation 1 - distribution of polygenic scores by group.

The resulting distributions are approximating a normal distribution (per the central limit theorem). Since we did not introduce any group differences in the assumptions, there are none to be seen aside from minor chance variation.

1.3.2. Simulation 2 – developmental noise

In the above, we implicitly assumed that polygenic genotype is identical to the phenotype (displayed variation in the organism). In real life however, no trait is exactly 100% heritable (though some come close), there is always some developmental noise. The amount of noise, as well as other factors not shared by siblings brought up together, collectively form the *unshared environment* variance component (E) estimated by classical twin studies. We can add this to our simulation by adding a small unshared environment component that consists of a standard normal distribution (mean of 0, standard deviation [sd] of 1). If we run the simulation with this change, the histogram will look like that in Figure 5. Note that we are now plotting the phenotypic score.

Figure 5: Simulation 2 - histogram of phenotypic scores by group.

Based on this simple assumptions, we have recreated the normal distribution so often seen in real data. However, the distributions have also become complex and not easy to visualize from the histogram. As such, we take another step and switch to using density fits. This consists of estimating the underlying (possibly non-normal) distribution of values and plotting the distribution instead of the data. The fitted distributions are shown in Figure 6.

Figure 6: Simulation 2 - distribution of phenotypic scores by group.

We can now more clearly see the underlying distributions. Because we introduced a noise component, the polygenic scores are no longer identical with the phenotypic scores and do not correlate perfectly with them. Figure 7 shows the correlations between the polygenic and phenotypic scores by group.

Figure 7: Simulation 2 - Scatterplot of polygenic and phenotypic scores by group.

The observed regression lines are of course identical by group, and the corresponding correlation is .85.

1.3.3. Simulation 3 – genetic group difference

So far we have been assuming that there are no non-chance group differences in their polygenic scores. However, in real life this is unlikely to be the case for all traits (examples will be discussed later). A group difference in polygenic scores consists of differences in the frequencies of one or more of the variants. To show this, we change the frequencies of the variants to .55 and .45 for group A and B, respectively. The resulting distributions are shown in Figure 8.

Figure 8: Simulation 3 - Distributions of phenotypic scores by group.

As expected, we now see a difference in the phenotypic scores as a result of the genetic difference we created. Unsurprisingly, if we calculate the mean polygenic scores by group, these are 5.5 and 4.5 for A and B respectively (number of variants times their probability by group). In terms of effect size, the phenotypic gap is 0.64 d (standardized) or 1 in raw units, representing a medium to large gap by the usual standards.

If we repeat the scatterplots that were done in simulation 3, we still find that the lines are identical, as seen in Figure 9.

Figure 9: Simulation 3 - Scatterplot of polygenic and phenotypic scores by group.

The absence of any difference in slope or intercept indicates that the group difference was completely due to genetic, or in other words, that the between group heritability was 100%. This is despite the fact that the within group heritability was not 100% (73%, i.e. the correlation of .86 squared).

If we model the data, we see that group membership becomes irrelevant once we take polygenic scores into account:

Model	Group membership	Polygenic score
1	1.00	
2	0.00	1.00

Table 2: Simulation 3 - Models of group differences. Unstandardized betas.

In other words, controlling for true cause, the non-causal correlate no longer had any validity.

1.3.4. Simulation 4 – mixed cause group difference

In reality life, group differences in many traits are unlikely to be entire due to genetics. Instead, the causal factors include both environmental and genetic causes. We can introduce this into our simulation by adding another normally distributed environmental factor (shared environment within group;

analogous to shared environment within sibling pairs in classical twin studies). Note that for it to contribute towards the group difference, the distributions must have different means by group. In our case, we use means of 1 and 0 for A and B, respectively. Figure 10 shows the phenotypic distributions.

Figure 10: Simulation 4 - Distribution of phenotypic scores by group.

The group difference has increased due to the addition of the environmental factor, and it now 2 in raw units. The standardized difference, however, remains at 0.64 d because the standard deviation (pooled) also increased.

Table 7 shows the correlations between the variables in the simulation.

	Polygenic score	Shared environment	Unshared environment	Phenotypic score
Polygenic score	1.00	0.14	0.00	0.77
Shared environment	0.14	1.00	0.01	0.58
Unshared environment	0.00	0.01	1.00	0.43
Phenotypic score	0.77	0.58	0.43	1.00

Table 3: Simulation 4 - Correlations between variables.

Interestingly, there is now a small positive correlation (.14) between polygenic score and shared environment, or in other words, there is a gene-environment correlation (rGE). This results due to the common cause, namely group membership.

Table 4 shows the results from the modeling, similar to before.

Model	Group membership	Polygenic score	Shared environment
1	2.00		
2a	1.00	1.00	
2b	1.00		1.01
3	0.00	1.00	1.00

Table 4: Simulation 4 - Models of group differences. Unstandardized betas.

It can be seen that leaving out either cause results in some of the validity to be switched to the group membership itself. Scatterplots of the relationships between the two causes and the phenotypic scores are shown in Figure 11.

Figure 11: Simulation 4 - Scatterplots of polygenic score/shared environment and phenotypic scores by group.

Unlike in simulation 3, there is now a difference in the intercepts of the regression lines, indicating some unmodeled cause. The slopes, however, are the same.

What is the between group heritability in simulation 4? Despite the unstandardized betas of the predictors being identical, the variance proportions (and correlations) are not the same. This is because the predictors do not have the same variance. Formally, the between group heritability is given by the equation below:

$$h_B^2 \approx h_W^2 \frac{(1-t)r}{(1-r)t}$$

where h^2 refers to the heritability, subscripts B and W to between and within groups, respectively, r is the genotypic intraclass correlation, and t is the phenotypic intraclass correlation.

The equation was worked out by DeFries (1972), and relies upon a number of approximations. If we apply it to the data from simulation 4, we find that the between group heritability is 25%, and the shared environmental part is thus 75% (since unshared environment does not contribute). This value cannot be obtained by simply squaring the correlation between polygenic score x phenotypic score in Table 3 because that would estimate the proportion of the total variance that is heritable (a mix of within and between group variance), not just the variance between groups.

1.3.5. Simulation 5 – Genomic ancestry and trait levels (admixture analysis)

In real life, we don't yet have accurate estimates of the true polygenic scores of persons. Thus, unless we can account for the measurement error in our modeling (Westfall & Yarkoni, 2016), we need to use another method. One such method is the admixture analysis (also called association studies; (Shriner, 2013)). Simulating data for this method is more complicated than the previous cases. For more reliable results, we need to increase the number of causal variants, and 100 was chosen as a reasonable compromise between computation time and realism. Second, we need to add non-causal locuses that are used to estimate the ancestry of a given person. For this, we add 1000 variants (also with just 0 and 1's). These are only used to score the ancestry of a person and don't affect the phenotype (i.e. we assume no pleiotropy). Third, we assume random mating, no sexes/hermaphrodites (everybody can mate with anybody), zero fatality (every one lives to breed), perfect recombination (recombination happens at every locus), and perfect generations (everybody reproduces at the same time with 2 offspring and then die). These assumptions are needed to simplify the mating process, but don't affect the results in a relevant way. To avoid very long computation times, we also decrease the group sizes to 5,000 for each. We then simulate our founder populations for each group, with the same situation as in simulation 3, 100% between group heritability combined with substantial but non-perfect within group heritability (genetics + noise).

To get admixed persons, we need to mate our total population. Since we assumed random mating, this is done by picking pairs at random until no one is left. Each pair then breeds 2 new individuals by picking a random half of their variants from each parent. This represents the next generation. We repeat this 5 times, so that we have 6 generations (0 through 5). The relationship between ancestry and phenotypic score by generation is shown in Figure 12.

Figure 12: Simulation 5 - Scatterplots of ancestry and phenotypic score for 6 generations.

There are two things of note here. First, that the predicted values at 0 and 100% ancestry stays the same despite the increasing admixture: they are always 45 and 55, which are the ancestry mean polygenic scores. Second, the correlations do not stay the same because they depend on the variation in ancestry. Since variation in ancestry decrease with each generation, the correlations must go down. The size of correlations (and standardized betas) cannot be used to estimate the importance of genetic causation for a given difference, one should use the model predicted values at 100% ancestry for a given group.

1.3.6. Simulation 6 – Non-random mating

Since, in simulation 5 we assumed random mating, it takes only a few generations to have no visible groups left. As such, there are no two fairly distinct groups we can calculate a between group heritability for, even though the genetic causes are still related to ancestry. Humans do not pick mates at random, and in fact, there is evidence that they prefer mates similar in ancestry, even within racial groups and when controlling for educational attainment (Helgason, Pálsson, Guðbjartsson, Kristjánsson, & Stefánsson, 2008; Zou et al., 2015). As a simple solution, mates were matched for ancestry for imperfectly such that some outbreeding would occur. The relationship between ancestry and phenotypic score is shown in Figure 13.

Figure 13: Simulation 6 - Scatterplots of ancestry and phenotypic score for 6 generations.

It can be seen that there is vastly more variation in ancestry in simulation 6 compared to 5. The standard deviations of ancestry in generation 5 (g5) were .34 and .09, for simulation 6 and 5, respectively. A simpler way to see this is to inspect the distribution of ancestry in each simulation side by side, as shown in Figure 14.

Figure 14: Comparison of ancestry distributions for simulation 5 (left) and 6 (right).

While in simulation the distribution is unimodal, it is still bimodal in simulation 6 after 5 generations of interbreeding.

With our bimodal ancestry distribution in hand, we can now simulate SIRE groups. This was done simply by using the ancestry as probability of being in group A, i.e. a person with 99% (A) ancestry would have a 99% chance of self-identifying as A. The results of this are shown in Figure 15.

Figure 15: Simulation 6 - Ancestry distribution by SIRE. Vertical lines show the mean.

As expected, each group was higher on its own ancestry and there was significant overlap between them.

Finally, we draw the regression lines by SIRE to see if we can correctly deduce the ancestry mean phenotypic scores. The scatterplot is shown in Figure 16.

Figure 16: Simulation 6 - Scatterplot of ancestry and Phenotypic score by SIRE.

As expected, the predicted values for 0% and 100% ancestry are the same for each SIRE group (45 and 55).

Finally, we apply the equation mentioned in Section 1.3.4. It correctly reached the result that 100% of the between group variance was genetic in origin.

At this point, one could go on introducing more complexities such as multiple ancestries, non-genetic causation (<100% between group heritability), but we think it is clear at this point that admixture analysis can be used to find indirect evidence of genetic causation. *Certainly, it can produce evidence incongruent with genetic models and thus offers a test of such.*

1.4. Cognitive and social inequality in the Americas

Taken as a whole, the Americas are populated almost entirely by three racial groups: Europeans, Africans and Amerindians.¹¹ Furthermore, there are significant numbers of mixed persons, especially in Southern America. While the racial classification systems differ by country, there is a fairly consistent ordering where European-derived SIRE groups tend to have higher cognitive ability and better social outcomes than African or Amerindian-derived SIREs, and with mixed groups in between (Fuerst &

11 A few areas have small numbers of Asians, e.g. West coast USA and Canada (various Asians), as well as Brazil (Brazilian Japanese, ~1% of total population). Generally, they are present in studies in too low numbers for accurate estimates and are usually ignored/excluded.

Kirkegaard, 2016b).

Researchers are divided with regards to the explanations of such differences, and as well as those present in the rest of the world (Rindermann et al., 2016). Some researchers argue that SIRE differences are solely the result of non-genetic differences such as education, income/wealth, access to technology, stereotype threat, discrimination and various other unspecified cultural differences (Colman, 2016; Nisbett, 2005, 2009; Nisbett et al., 2012; Sternberg, 2005; Suzuki & Aronson, 2005). One prominent behavioral geneticist even claims to be able to derive the lack of genetic causation from ethical principles (Turkheimer, 2007).¹² Other researchers argue that such differences are better explained by some combination of genetic and environmental causes (Jensen, 1973; Loehlin, Lindzey, & Spuhler, 1975; Herrnstein & Murray, 1994; Jensen, 1998; Rushton, 2000; Rushton & Jensen, 2005; Lynn, 2015).

What evidence exists regarding causation? It is not our goal here to extensively review the literature, but some previous studies need to be summarized. Before the advent of genetic testing, genomic ancestry was estimated by relying on visual traits like skin tone, as well as blood groups and family history (Loehlin et al., 1975). While these methods work to some degree, they introduce substantial measurement error into the analysis (Shriver et al., 2003) which thus complicates interpretation. Perhaps unsurprisingly, then, there is disagreement about whether the existing, published evidence is more congruent with a relationship between genomic ancestry and cognitive ability or not (Colman, 2016; Loehlin et al., 1975; Lynn, 2002, 2015; Nisbett, 2009; Rushton & Jensen, 2005).

All previous studies relied on various proxies or poor measures of genomic ancestry and the samples were usually small, and together this probably explains the inconsistent results.¹³ As such, no published study has examined genomic ancestry in a large sample and related this to cognitive ability. However, a large meta-analysis of genomic ancestry and socioeconomic outcomes¹⁴ in American countries found that European ancestry was robustly associated with better outcomes ($r = .18$, $k = 27$, total $n = 34k$; (Kirkegaard, Wang, & Fuerst, 2017)). Reversely, Amerindian and African ancestry were related to worse outcomes (r 's of $-.15$ and $-.11$, respectively¹⁵). Because of the moderate to strong causal relationship between cognitive ability and socioeconomic outcomes (Gottfredson, 1997; Herrnstein &

12 “[...] Why don’t we accept racial stereotypes as reasonable hypotheses, okay to consider until they have been scientifically proven false? They are offensive precisely because they violate our intuition about the balance between innateness and self-determination of the moral and cultural qualities of human beings. No reasonable person would be offended by the observation that African people have curlier hair than the Chinese, notwithstanding the possibility of some future environment in which it is no longer true. But we can recognize a contention that Chinese people are genetically predisposed to be better table tennis players than Africans as silly, and the contention that they are smarter than Africans as ugly, *because it is a matter of ethical principle that individual and cultural accomplishment is not tied to the genes in the same way as the appearance of our hair.*” (our emphasis)

13 Nisbett, for instance, gives heavy weight to a 1986 study involving 46 7-10 year old children (Moore, 1986) and a 1972 study involving 85 2-5 year old children (Tizard, Cooperman, Joseph, & Tizard, 1972). If the replication crises has taught us anything, it is not trust old, small, unreplicated studies, especially when there is a very strong social pressure to publish results in one direction (Cofnas, 2015; Duarte et al., 2015; Flore & Wicherts, 2015; Open Science Collaboration, 2015).

14 Educational attainment, income, wealth, occupational status/prestige etc.

Murray, 1994; Strenze, 2007), it is unlikely that European ancestry would not be associated with higher cognitive ability.

The purpose of this study was to examine the relationship between cognitive ability, socioeconomic outcomes and genomic ancestry in a large sample.

2. Data and coding of variables

We utilized data from the PING (Pediatric Imaging, Neurocognition, and Genetics; (Akshoomoff et al., 2014)) dataset. PING is a neuroscientific dataset based on a large sample ($n = 1,493$) of healthy American children and youths, ages 3 to 21, of diverse ancestries. The data has been employed in a number of studies to explore, among other things, the relationship between brain morphology and genomic ancestry (Fan et al., 2015; Noble et al., 2015). Participants were not nationally representative but were rather recruited from greater metropolitan areas of Baltimore, Boston, Honolulu, Los Angeles, New Haven, New York, Sacramento, and San Diego. Individuals which had medical conditions which could affect development were excluded from the recruitment process.

Since we expected some of the ancestry effects to be fairly small relative to the total sample size in this study, we generally made analytic decisions that maximized our sample size as opposed to maximizing the data quality.

All analyses were done in R (a statistical computing language). Source code and detailed output from the analyses as well as many other figures and tables are available in the supplementary materials.

2.1. Cognitive data

Participants were given the NIH toolbox cognitive tests. The cognitive sub-domain assessed by each test, along with the subtests' psychometric characteristics, are detailed by Akshoomoff et al (2014). The seven cognitive tests were as follows: Dimensional Change Card Sort Test (Card Sort); Flanker Inhibitory Control and Attention Test (Flanker); Picture Sequence Memory Test (Picture Sequence); Pattern Comparison Processing Speed Test (Pattern Recognition); Oral Reading Recognition Test (Pattern Recognition); List Sorting Working Memory Test (List Sort); and Picture Vocabulary Test (Vocabulary).

Some data were missing (5% of cells, 16% of cases had at least 1 missing datapoint). We imputed values for cases that were missing less than half of their datapoints rounded down (3 or fewer). Imputation was done using IRMI (Templ, Alfons, Kowarik, & Prantner, 2015; Templ, Kowarik, & Filzmoser, 2011). After this, there were 1,452 cases with complete data.

Next, we regressed out the effect of sex and age because these were not of interest in this study. This was done by first fitting a local regression model (LOESS; James, Witten, Hastie, & Tibshirani, 2013)

15 The relatively smaller size of these correlations should not be strongly interpreted because the size of the correlation is dependent on the amount of variation in the ancestry in question, and it's relative difference in comparison to other ancestry present in the sample.

with age as the predictor and saving the residuals. These were then residualized for sex using a standard linear model (OLS). We used LOESS because this approach is able to capture complex non-linear patterns in the data.

Finally, we factor analyzed the adjusted data for the 7 tests and extracted a g factor (Jensen, 1998). The pattern of factor loadings looked normal.¹⁶ As shown in Table 5, tests exhibited the expected positive manifold. Scores on this factor were saved and standardized (mean = 0, sd = 1) for further analysis. We were faced with the choice of whether to standardize the scores to the White subsample or whether to use the entire sample. We chose to use the full sample because this would likely counterbalance the likely cognitive ability/social status selection of the sample due to urban sampling.

	Picture sequence	List sort	Pattern recognition	Vocabulary	Reading	Flanker	Card sort
Picture sequence	1.00	0.34	0.13	0.22	0.20	0.10	0.16
List sort	0.34	1.00	0.18	0.34	0.31	0.22	0.30
Pattern recognition	0.13	0.18	1.00	0.13	0.13	0.29	0.32
Vocabulary	0.22	0.34	0.13	1.00	0.56	0.17	0.22
Reading	0.20	0.31	0.13	0.56	1.00	0.17	0.21
Flanker	0.10	0.22	0.29	0.17	0.17	1.00	0.52
Card sort	0.16	0.30	0.32	0.22	0.21	0.52	1.00

Table 5: Cognitive test intercorrelations after adjusting for age and sex.

In general, the intercorrelations are on the low side. This perhaps has to do with residual age confounding in the data given that many abilities/tests show non-linear trends throughout childhood development.

2.2. Parental socioeconomic data

Because the subjects in this dataset were children, they had no socioeconomic outcomes. However, the following parental variables were available:

- Household income
- Parent 1 educational level
- Parent 2 educational level
- Parent 1 occupational level

¹⁶ Loadings were, in order: vocabulary .60, reading .57, list sort .57, card sort .54, flanker .47, picture sequence .38, pattern recognition .36.

- Parent 2 occupational level

The familial participant-guardian relationships were reported. For 95% of the participants for which relationship data were available, guardian 1 was either a biological mother or father (missing data: 2%; self: 1%) and for 92% of the participants for which relationship data were available, guardian 2 was either a biological mother or father (missing data: 32%). Since most often guardians were parents, we refer to guardians as parents. As before, some data were missing. Since we were interested in maximizing the available sample size, we imputed all data with at least one datapoint. Finally, we factor analyzed the variables to extract a general socioeconomic factor (S; Kirkegaard, 2014; Kirkegaard & Fuerst, 2017). Scores on this factor were saved and standardized (mean = 0, sd = 1) for further analysis. As before, the standardization was done on the full sample, rather than the White subset.

2.3. Demographic data

Genomic ancestry estimates were available in the PING data files. These ancestry estimates were calculated using the ADMIXTURE algorithm based on more than 16k SNPs and are thus very likely to be highly reliable (see for details, Akshoomoff et al., 2014). Ancestry was assigned to 6 clusters corresponding to Europeans, Africans, Amerindian (Native Americans), East Asians, Oceanians and Central Asians. All ancestry was forced to be assigned to one of these clusters and because of this, the ancestry variables for each person summed to 1.00. Ancestry data were available for 1,391 persons.

The SIRE data presented us with the choice of how they should be coded. Subjects were asked for each of a set of 6 SIREs whether they considered it to apply to themselves or not. These were: Hispanic/Latino, Pacific Islander, Asian, African American, American Indian, and White. We coded the data in four different ways. This was done to assess to which degree the coding decision influenced the results (researcher degrees of freedom/method variance: Gelman & Loken, 2013; Silberzahn et al., In review; Simmons, Nelson, & Simonsohn, 2011; Wicherts et al., 2016).

Some cases had missing data (2% of cells, 2% of cases had at least one missing datapoint). We filled in missing data with FALSE.

2.3.1. Coding method 1 – standard coding

The first coding method followed standard US practice of giving special status to Hispanic ethnicity. The approach was:

1. Does the person self-identify as Hispanic? If yes, code as Hispanic.
2. Does the person self-identify as a single of the non-Hispanic SIREs? If yes, code as that SIRE.
3. Does the person self-identify as multiple of the non-Hispanic SIREs? If yes, code as Multi-ethnic.
4. Does the person self-identify as none of the options? Code as Other.

The breakdown of SIREs is shown in Table 6.

Group	Count	Percent
White	571.00	41.05
Hispanic	328.00	23.58
Multi-ethnic	186.00	13.37
African American	140.00	10.06
Asian	122.00	8.77
Other	24.00	1.73
Pacific Islander	16.00	1.15
American Indian	4.00	0.29

Table 6: SIRE groups – standard coding.

2.3.2. Coding method 2 – common combinations coding

Because it was possible for subjects to select multiple SIREs, and because different combinations of multiple SIREs may be quite different (e.g., White + African American vs. Asian + American Indian), for coding method 2 we coded each person as the conjunction of all their chosen SIREs, giving no special status to Hispanic ethnicity. These codings, as in coding method 1, were mutually exclusive and exhaustive. This produced a total of 47 unique combinations. Most of these had very small sample sizes and would thus not be informative in modeling (e.g., 8 combinations had only 1 person). As such, we grouped categories with fewer than 29 persons – a somewhat arbitrarily selected number – into a remainder category (“Other”). The resulting 9 groups are shown in Table 3. The resulting 9 groups are shown in Table 7.

Group	Count	Percent
White	571.00	41.05
Other	226.00	16.25
African American	140.00	10.06
Hispanic, White	140.00	10.06
Asian	122.00	8.77
Hispanic	71.00	5.10
Asian, White	60.00	4.31

Pacific Islander, Asian, White	32.00	2.30
Hispanic, African American	29.00	2.08

Table 7: SIRE groups – common combinations coding.

2.3.3. Coding method 3 – continuous coding

As discussed above, it is problematic to group all persons who select more than one SIRE together. However, creating categories for all possible combinations, as above, produces many small groups that are not useful for analysis. One solution to the analytic problem presented by multiple-SIRE individuals is to code the participants' SIRE as an interval or continuous variable (Liebler & Halpern-Manners, 2008). In cultural terms, this makes sense, as a person who self-identifies as White and African American can be said to belong “half” to each SIRE culture. This coding was done by first counting the number of selected SIREs and then assigning 1 divided by that number to each of the selected SIREs. For example, a person who chose White, Asian, and Hispanic would be assigned scores of 1/3 to each of these categories and 0 to the remaining ones.

2.3.4. Coding method 4 – dummy coding

Finally, because of the lack strict interdependence between the SIRE variables in the dataset, one can also enter each variable as a dummy (dichotomous categorical) predictor. The difference between this method and combinations coding is that coding as dummy variables captures only main effects associated with each SIRE, while coding for combinations also captures interaction effects at the cost of reduced precision (bias-variance tradeoff; James et al., 2013).

2.4. Other variables

There were three almost perfectly correlated age variables in the dataset. Since we wished to control for the effect of age on raw test scores, we used age at neuropsychological testing as the primary variable and filled in missing data ($n = 3$) with age at MRI. With this strategy, age data were available for 1,386 cases. The mean age was 12 with a standard deviation of 5 and a range of 3-21.

Because cognitive tests are often thought to be biased against non-native speakers, we also included a dichotomous English as First Language (EFL) variable as a control in the regression analyses. For 11% of the sample, English was not the first language.

3. Analyses

3.1. SIRE and genomic ancestry

It is known that genomic ancestry varies substantially within some SIREs. African Americans average

about 80% African ancestry and 20% European (Bryc et al., 2015; Shriver et al., 2003). There is substantial variation with a few persons having nearly 100% European ancestry. To examine the distribution of genomic ancestry by SIRE in the current dataset, we calculated mean and standard deviation of genomic ancestries by SIRE (standard coding), shown in Tables 8-9.

SIRE	n	European	African	Amerindian	East Asian	Oceanian	Central Asian
White	571	0.97	0.00	0.01	0.00	0.00	0.01
African American	140	0.17	0.81	0.00	0.00	0.00	0.01
American Indian	4	0.66	0.21	0.13	0.00	0.01	0.00
Asian	122	0.06	0.00	0.00	0.80	0.01	0.13
Hispanic	328	0.57	0.14	0.19	0.09	0.01	0.01
Multi-ethnic	186	0.45	0.11	0.01	0.37	0.03	0.03
Other	24	0.55	0.26	0.01	0.01	0.00	0.18
Pacific Islander	16	0.11	0.01	0.00	0.70	0.17	0.00

Table 8: Mean genomic ancestry by SIRE (standard coding).

SIRE	n	European	African	Amerindian	East Asian	Oceanian	Central Asian
White	571	0.07	0.04	0.02	0.04	0.00	0.04
African American	140	0.11	0.12	0.01	0.03	0.01	0.02
American Indian	4	0.42	0.42	0.25	0.00	0.01	0.00
Asian	122	0.17	0.02	0.01	0.36	0.01	0.32
Hispanic	328	0.23	0.20	0.17	0.20	0.02	0.03
Multi-ethnic	186	0.27	0.21	0.02	0.29	0.05	0.12
Other	24	0.36	0.33	0.03	0.02	0.01	0.32
Pacific Islander	16	0.15	0.04	0.00	0.13	0.06	0.00

Table 9: Standard deviation of genomic ancestry by SIRE (standard coding).

As expected, Whites were almost entirely genomically European (97%) as has been found in previous studies of US citizens (Bryc et al., 2015; Shriver et al., 2003), while the other groups showed more mixed ancestry.

We also looked at the correlations between continuous SIRE and genomic ancestry, which are shown in

Table 10. It can be seen that in the US, continuous SIRE is a reasonably good index of genomic ancestry. This is because major US SIRE groups are fairly unadmixed (as compared to, e.g., Brazilian ones), and thus much of the variance is explained by recent patterns of exogamy which can be captured by the type of coding we use here.

SIRE	n	European	African	Amerindian	East Asian	Oceanian	Central Asian
White	723	<i>0.86</i>	-0.50	-0.23	-0.44	-0.22	-0.13
African American	179	-0.45	<i>0.92</i>	-0.14	-0.18	-0.08	-0.06
Hispanic	187	-0.08	0.00	<i>0.75</i>	-0.15	-0.06	-0.08
American Indian	27	0.00	0.03	<i>0.09</i>	-0.04	-0.01	-0.04
Asian	196	-0.56	-0.20	-0.16	<i>0.81</i>	0.09	0.27
Pacific Islander	55	-0.28	-0.10	-0.07	0.39	<i>0.81</i>	-0.03
Other	24	-0.03	0.06	-0.04	-0.06	-0.03	<i>0.16</i>

Table 10: Correlations between genomic ancestry and self-identified race/ethnicity (SIRE, continuous coding). Theoretical expected correlations marked in italics. Note that n's here refer to complete persons for each SIRE (i.e. summing across all cases).

In general, correlations between the matching SIRE and genomic ancestry pairs were strong. The few exceptions to this (e.g. Amerindian x American Indian) is due to very small number of persons with this SIRE in relationship to the competing SIRE for the same genomic ancestry cluster (Hispanic).

3.2. Bivariate relationship between genomic ancestry, cognitive ability and parental social status

The correlations between cognitive ability (CA), parental general socioeconomic status (S) and each of the genomic ancestries are shown in Table 11.

	CA	S	European	African	Amerindian	East Asian	Oceanian	Central Asian
CA	1.00	0.42	0.23	-0.33	-0.15	0.05	-0.08	0.06
S	0.42	1.00	0.32	-0.30	-0.24	-0.06	-0.20	0.12
European	0.23	0.32	1.00	-0.50	-0.10	-0.62	-0.31	-0.20
African	-0.33	-0.30	-0.50	1.00	-0.08	-0.22	-0.10	-0.07
Amerindian	-0.15	-0.24	-0.10	-0.08	1.00	-0.16	-0.09	-0.07
East Asian	0.05	-0.06	-0.62	-0.22	-0.16	1.00	0.42	-0.07

Oceanian	-0.08	-0.20	-0.31	-0.10	-0.09	0.42	1.00	-0.03
Central Asian	0.06	0.12	-0.20	-0.07	-0.07	-0.07	-0.03	1.00

Table 11: Correlations between cognitive ability (CA), parental general socioeconomic status (S) and each of the genomically measured ancestries. $n = 1369$ -1391.

The correlations were roughly as would be expected given the well-documented cognitive and socioeconomic differences between US SIRE groups. European ancestry had a moderate positive correlation with both cognitive ability and SES ($r = .23$ and $.32$, respectively), while negative relationships were seen for African ($r = -.33$ and $-.30$), Amerindian ($r = -.15$ and $-.24$), and Oceanian ($r = -.08$ and $-.20$) ancestries. The genomic ancestry $\times$ S correlations are larger than those typically reported by Kirkegaard et al. (2017), presumably because this sample is not decomposed by SIRE group, leading to expanded range and hence higher correlations. (See the supplementary materials for correlation matrices by subpopulations.) The remaining ancestries had very small or inconsistent relationships.

Previous research has found a modest relation between parental SES and cognitive ability (Sirin, 2005). In our sample, the relation ($r = .42$) was somewhat stronger than typically reported. This increased strength of association is probably in part due to the use of a general socioeconomic factor, which is a more reliable measure, as opposed to individual measures of parental socioeconomic status.

Not much should be made of the bivariate analysis for ancestry and outcomes because these confound the effects of multiple ancestries as well as the effects of non-genetic causes. In particular, the ancestry variables are necessarily negatively related in general because every individual's ancestry must sum to 1.

3.3. Main regression analyses

Our main regression analysis strategy consisted of fitting models with the genomic ancestries and one of the SIRE coding variants. Note that we left out European ancestry to avoid perfect linear dependency among the predictors. As such, the betas of the included ancestries should be seen relative to European ancestry.¹⁷ This was also done for the continuous SIRE coding for the same reason. Because we have 4 SIRE coding variants, a dichotomous choice about whether to include the genomic ancestry predictors, and 2 outcomes, this results in 16 models. Because the results were quite similar, we only report the detailed results for SIRE with standard coding here (the remainder can be found in the supplementary materials). We choice this coding for reader familiarity. The results from the models are shown in Tables 12 and 13. Note that we report unstandardized betas in line with our comments about the use of standardized betas/correlations in Section 1.3.

	Model 1, adj. $r^2 = .11$	Model 2, adj. $r^2 = .16$
--	---	---

¹⁷ The same approach was used previously (Fuerst & Kirkegaard, 2016a).

Predictor	Beta	SE	CI lower	CI upper	Beta	SE	CI lower	CI upper
SIRE standard: White	0.00				0.00			
SIRE standard: African American	-0.97	0.09	-1.15	-0.79	0.31	0.18	-0.05	0.67
SIRE standard: American Indian	-0.09	0.47	-1.02	0.84	0.41	0.46	-0.50	1.32
SIRE standard: Asian	-0.02	0.10	-0.21	0.17	-0.07	0.19	-0.44	0.30
SIRE standard: Hispanic	-0.62	0.07	-0.75	-0.48	-0.16	0.10	-0.36	0.03
SIRE standard: Multi ethnic	-0.24	0.08	-0.40	-0.08	0.04	0.11	-0.18	0.25
SIRE standard: Other	-0.33	0.22	-0.76	0.10	-0.03	0.23	-0.48	0.42
SIRE standard: Pacific Islander	-0.76	0.24	-1.23	-0.29	0.02	0.31	-0.58	0.63
African					-1.58	0.20	-1.97	-1.19
Amerindian					-1.15	0.31	-1.76	-0.55
East Asian					0.06	0.19	-0.31	0.42
Oceanian					-4.81	1.27	-7.31	-2.31
Central Asian					0.21	0.26	-0.29	0.71
English first language	0.04	0.09	-0.13	0.21	0.03	0.08	-0.14	0.20

Table 12: Regression model results. Outcome: cognitive ability. Standard SIRE coding. n = 1369.

	Model 1, adj. r² = .17				Model 2, adj. r² = .24			
Predictor	Beta	SE	CI lower	CI upper	Beta	SE	CI lower	CI upper
SIRE standard: White	0.00				0.00			
SIRE standard: African American	-0.98	0.09	-1.14	-0.81	0.28	0.17	-0.06	0.63
SIRE standard: American Indian	-1.45	0.46	-2.34	-0.55	-0.85	0.44	-1.71	0.02
SIRE standard: Asian	-0.17	0.09	-0.36	0.01	-0.03	0.18	-0.38	0.32
SIRE standard: Hispanic	-0.84	0.06	-0.97	-0.71	-0.21	0.09	-0.39	-0.03
SIRE standard: Multi ethnic	-0.46	0.08	-0.61	-0.31	-0.03	0.10	-0.23	0.17
SIRE standard: Other	-0.52	0.24	-0.99	-0.05	-0.39	0.24	-0.85	0.08
SIRE standard: Pacific Islander	-1.62	0.23	-2.07	-1.17	-0.22	0.29	-0.79	0.35

African					-1.55	0.19	-1.93	-1.18
Amerindian					-1.91	0.29	-2.48	-1.34
East Asian					-0.24	0.18	-0.58	0.11
Oceanian					-7.23	1.20	-9.59	-4.87
Central Asian					0.62	0.25	0.14	1.11
English first language	0.13	0.08	-0.03	0.29	0.11	0.08	-0.05	0.27

Table 13: Regression model results. Outcome: parental general socioeconomic factor. n = 1380.

Initially it should be noted that that some groups and ancestries had very small samples, and results for them should be disregarded (e.g. American Indian, n = 4, cf. Table 6). We included all the groups into the models to avoid decreasing our sample size, as excluding them would have done.

In general, the two models show convergent results: before inclusion of the genomic ancestries, the SIREs had strong validity. For instance, African American status SIRE had a beta of -0.97 for cognitive ability, meaning that this group was -0.97 d below the sample average. However, once genomic ancestry was included in the model, SIREs lost most of their detectable validity, and sometimes changed sign. For cognitive ability, 13 of the 14 SIRE betas had confidence intervals overlapping 0, and the last only barely (Hispanic for parental S).

With regards to ancestry, large betas can be seen. For African ancestry, betas are about -1.60 meaning that a person with 100% African ancestry, compared with one with 100% European, is predicted to be 1.6 d below a person with 0% African ancestry on both cognitive ability and parental S. For Amerindian ancestry the same number the betas differed by outcome, being -1.15 for cognitive ability and -1.91 for parental S. This might reflect the fact that the parents of many of the children with high Amerindian ancestry were recent immigrants still low in social status, while their children benefited from the improved US environment. Improvements in cognitive scores from 1st to 2nd generation immigrants, both Hispanic and others, have long been noted (Dunn, 1988; Richwine, 2009; Rindermann & Thompson, 2014). East Asian ancestry did not seem to have any detectable influence on either outcome, while Central Asian only had a marginally detectable effect on parental S.

English as first language did not have any detectable effects on either outcomes, whether or not in was included along with genomic ancestry.

3.3.1. Aggregate results across SIRE codings

Above we mentioned that the results were similar across the 4 codings of SIRE, but how similar? Unfortunately, it is not so easy to compare all the predictor estimates directly because the SIRE codings result in different numbers of and often unpairable betas per model.¹⁸ However, the ancestry estimates

18 For instance, common combinations results in the SIRE group 'Hispanic, White' which has no clear counterpart in standard coding (note n's for standard Hispanic = 328, but common combinations Hispanic+White = 140).

remain constant across models and are thus easy to compare. Table 14 shows the summary statistics for the ancestry betas across models.

Predictor	Cognitive ability mean (min, max)	Parental S mean (min, max)
African	-1.46 (-1.59 to -1.31)	-1.47 (-1.76 to -1.23)
Amerindian	-0.92 (-1.15 to -0.75)	-1.90 (-2.06 to -1.76)
Central Asian	0.17 (0.15 to 0.21)	0.81 (0.62 to 1.07)
East Asian	0.03 (-0.01 to 0.06)	0.02 (-0.24 to 0.33)
Oceanian	-2.76 (-4.81 to -1.32)	-4.02 (-7.23 to -1.58)

Table 14: Summary statistics of genomic ancestry betas across models.

It can be seen that the different SIRE codings did not strongly affect the betas of the ancestries with the exception of Oceanian, which as mentioned earlier had extremely small sample size. Overall, the betas for cognitive ability and parental S were similar, but there were some patterns. Amerindian ancestry consistently predicted much worse social outcomes than cognitive ability, consistent with the hypothesis that Hispanic S is suppressed due to recent immigration status (Borjas, 2016). On the other hand, Central Asian ancestry predicted better social outcomes than cognitive outcomes, perhaps due to higher non-cognitive traits for this ancestry that boost S beyond what one would expect based on their cognitive ability.

Which SIRE coding produced the best model fits? Table 15 shows the model fit (r^2 adj.) by SIRE coding and outcome.

SIRE coding	r² adj.	Outcome
standard	0.16	CA
common	0.17	CA
numeric	0.17	CA
dummy	0.17	CA
standard	0.24	S
common	0.25	S
numeric	0.27	S
dummy	0.28	S

Table 15: Model fit by SIRE coding and outcome.

The differences between coding were not large, particularly not for cognitive ability. However, there was a tendency for dummy coding to work best.

3.3.2. Parental S as a predictor

Parental S could be seen as a confound in the models where children's cognitive ability is the outcome. It is known that shared environment, which includes parental S, affects children's cognitive ability at the age of the children in this study (mean age 12, sd 5, range 3-21; (Bouchard, 2013)). On the other hand, if genomic ancestry is related to causal variants for cognitive ability and parental S, controlling for parental S indirectly controls for variation in children's genomic ancestry and as such constitutes an over-correction, a variant of the sociologist's fallacy (Jensen, 1973, p. 235; Rowe & Rodgers, 1997; Sesardić, 2005, p. 121).¹⁹ Figure 17 shows a theoretical path model of the variables.

¹⁹ This is also true if genomic ancestry has no causal effect, but is a proxy for some other variable (unmeasured familial variables) that does have a causal effect.

Figure 17: Theoretical path model for relationships between genomic ancestry, cognitive ability and parental S.

We did not have measures of parental variables aside from S for the present dataset and we are thus unable to fit a path model. Instead, we report the regression results for what we have but caution interpretation of them. Results for standard SIRE are shown in Table 16 (the 3 others are similar and can be found in the in supplementary materials).

	Model 1, adj. r ² = .21				Model 2, adj. r ² = .23			
Predictor	Beta	SE	CI lower	CI upper	Beta	SE	CI lower	CI upper
Parental S	0.35	0.03	0.29	0.40	0.31	0.03	0.26	0.36
SIRE standard: White	0.00				0.00			
SIRE standard: African American	-0.62	0.09	-0.80	-0.45	0.19	0.18	-0.16	0.54
SIRE standard: American Indian	0.42	0.45	-0.46	1.29	0.66	0.44	-0.21	1.53
SIRE standard: Asian	0.05	0.09	-0.13	0.23	-0.08	0.18	-0.43	0.28
SIRE standard: Hispanic	-0.32	0.07	-0.45	-0.19	-0.11	0.10	-0.30	0.08
SIRE standard: Multi ethnic	-0.08	0.08	-0.23	0.07	0.03	0.11	-0.17	0.24
SIRE standard: Other	-0.01	0.23	-0.47	0.45	0.14	0.24	-0.33	0.61
SIRE standard: Pacific Islander	-0.20	0.23	-0.64	0.25	0.09	0.29	-0.49	0.66
Ancestry: African					-1.05	0.20	-1.44	-0.66

Ancestry: Amerindian					-0.53	0.30	-1.12	0.05
Ancestry: East Asian					0.16	0.18	-0.20	0.51
Ancestry: Oceanian					-2.62	1.23	-5.04	-0.20
Ancestry: Central Asian					0.04	0.25	-0.45	0.53
English first language	0.00	0.08	-0.16	0.15	0.00	0.08	-0.16	0.16

Table 16: Regression model results. Outcome: cognitive ability. n=1363.

In the model without genomic ancestry, we see that the beta from African American and Hispanic SIRE have been markedly reduced (-0.97 to -0.62 (36%) and -0.62 to -0.32 (48%), respectively) as one would expect when conditional on average parental S. However, once genomic ancestry is added to the model, the betas for African American and Hispanic SIREs are reduced to near-zero, and the betas for African and Amerindian ancestry remain far from zero. Though it should be noted that the Amerindian beta's confidence interval overlaps with zero, due to low precision (standard error = .30).

Interpretation of the results is unclear as the reduction in validity of ancestry predictors is expected on both genetic and non-genetic models. For non-genetic models, they are expected due to controlling for the parental S causal confounder, while for genetic models they are expected due to indirectly controlling for parental cognitive ability through the control for parental S (Jensen, 1969, 1973, Chapter 11; Sesardić, 2005, p. 121).

3.4. Subgroup analyses

As a robustness check, we carried out a number of subgroup analyses to check whether results held when data were analyzed in specific subgroups. For this purpose, we chose three subgroups of interest: 1) non-Whites, 2) African Americans, and 3) Hispanics.

3.4.1. Non-Whites

For this analysis, we excluded everyone who selected only White as their race, i.e. non-Hispanic White in the standard SIRE coding. For SIRE predictors, we used dummy coding which enables the inclusion of White SIRE, though in this case it refers to people who choose White as one of their multiple SIREs, not people who only chose White. Tables 17 shows the regression results.

	Model 1, cognitive ability, adj. r² = .18, n=802				Model 2, parental S, adj. r² = .25, n=810			
Predictor	Beta	SE	CI lower	CI upper	Beta	SE	CI lower	CI upper
SIRE dummy: White	0.20	0.09	0.02	0.38	0.28	0.09	0.11	0.45
SIRE dummy: African American	0.30	0.16	-0.02	0.61	0.28	0.16	-0.03	0.59
SIRE dummy: American	-0.12	0.13	-0.36	0.13	-0.21	0.12	-0.45	0.02

Indian								
SIRE dummy: Asian	0.20	0.13	-0.06	0.46	0.04	0.13	-0.21	0.29
SIRE dummy: Hispanic	-0.20	0.10	-0.40	0.00	-0.16	0.10	-0.35	0.03
SIRE dummy: Other	0.16	0.25	-0.33	0.64	-0.17	0.26	-0.69	0.34
SIRE dummy: Pacific Islander	-0.45	0.14	-0.72	-0.18	-1.00	0.13	-1.25	-0.74
Ancestry: African	-1.47	0.26	-1.98	-0.96	-1.40	0.26	-1.91	-0.90
Ancestry: Amerindian	-0.81	0.34	-1.48	-0.15	-1.68	0.33	-2.32	-1.04
Ancestry: East Asian	0.00	0.21	-0.41	0.41	0.02	0.20	-0.38	0.42
Ancestry: Oceanian	-1.34	1.37	-4.02	1.34	-1.53	1.31	-4.10	1.04
Ancestry: Central Asian	0.10	0.27	-0.44	0.64	0.70	0.27	0.17	1.23
English first language	0.09	0.10	-0.10	0.29	0.28	0.09	0.10	0.46

Table 17: Regression results for non-Whites.

The results are similar to the full sample analyses in that SIREs generally have borderline detectable validity while African and Amerindian ancestries retained strong betas.

3.4.2. African Americans

We included only those who self-identified as African Americans and no other SIRE corresponding to the standard doing of African American. We did not include SIRE predictors for these regressions as we only had one SIRE. Results are shown in Table 18. Because African Americans only have 2 ancestry components (African and European sum to 98%) it is possible to easily display the regression results, which are shown in Figure 18, similar to that from the simulation data (Figure 16). Whites are included for ease of comparison. Note that one could have plotted European ancestry instead, but this would accomplish nearly the same due to the near -1 (-.93) negative correlation between the two ancestries.

	Model 1a, cognitive ability, adj. r ² = .04, n=138				Model 2a, parental S, adj. r ² = .09, n=139			
Predictor	Beta	SE	CI lower	CI upper	Beta	SE	CI lower	CI upper
Ancestry: African	-1.60	0.61	-2.81	-0.39	-2.50	0.65	-3.79	-1.22
	Model 1b, cognitive ability, adj. r ² = .02, n=138				Model 2b, parental S, adj. r ² = .07, n=139			
Ancestry: European	1.36	0.66	0.06	2.66	2.28	0.70	0.89	3.66

Table 18: Regression results for African Americans.

Figure 18: Scatterplot of European ancestry and cognitive ability in African Americans and Whites (standard SIRE). $n=140$.

Comparing the estimated cognitive ability at 0% and 100% African ancestry gives the estimate of the genetic gap, *assuming a simple admixture model*. This gap comes out to -1.60 or approximately 24 IQ (76 relative to European of 100).

The results for parental S were stronger the gap estimated at -2.50. The somewhat varying size estimates across the two outcomes should not be taken too seriously because of the strong uncertainty on the betas in this small subsample (standard errors of approx. 0.65).

3.4.3. Hispanics

For Hispanics, we included everybody who self-identified as Hispanic. With respect to which ancestries to include in the models, we included all ancestries with a mean prevalence of $\geq 5\%$, which resulted in the inclusion of European, African, Amerindian and East Asian (cf. Table 8). As per our usual approach, we left out European ancestry to avoid (near-)perfect collinearity. The regression results are shown in Table 19.

	Model 1, cognitive ability, adj. r2 = .08, n=323				Model 2, parental S, adj. r2 = .19, n=324			
Predictor	Beta	SE	CI lower	CI upper	Beta	SE	CI lower	CI upper
Ancestry: African	-1.62	0.30	-2.22	-1.03	-2.00	0.26	-2.52	-1.48
Ancestry: Amerindian	-1.16	0.34	-1.84	-0.49	-2.09	0.30	-2.69	-1.49
Ancestry: East Asian	-0.53	0.28	-1.09	0.02	-1.32	0.25	-1.81	-0.82

Table 19: Regression results for Hispanics.

Interestingly, the estimates for African and Amerindian ancestry were similar to those found in the African American subsample. This is consistent with simulations (see supplementary materials) showing that admixture analysis is not biased due to the presence of >2 ancestries as long as they are modeled appropriately. The results for East Asian ancestry were not in line with those found in the main or non-White analyses. This finding is puzzling and may be due to heterogeneity within the East Asian ancestry such that the subset of East Asians who mate with African Americans and Hispanics tend to be lower in cognitive ability and S (e.g. Cambodians or Philippine not Japanese or Koreans). The matter could be further investigated if one estimated sub-East Asian ancestries, as has been done extensively for European ancestries (Bryc et al., 2015; see also Kirkegaard, 2016a).

4. Discussion and conclusion

We examined a fairly large, ancestrally diverse dataset and found genomic ancestry to be related to both cognitive ability and parental general socioeconomic status (S) in ways that genetic models of racial inequality predict, i.e. that African and Amerindian ancestries predicted lower cognitive ability and S in comparison with European ancestry (the reference). The validity of self-identified race/ethnicity (SIRE) memberships were generally not detected with certainty (i.e. p values generally > .05, no correction for multiple testing). Larger samples are needed to clarify the predictive validity of SIRE because of the very strong associations with genomic ancestry (cf. Table 8).

We did not have socioeconomic outcomes for the subjects, only for their parents. This is not a major problem for interpretation of the results because variation in parental S cannot cause variation in genomic ancestry in their children; the genomic ancestry of the children is solely caused by parental genomic ancestry and random events (recombination). This means that using children's ancestry is essentially a very good proxy for the mean parental ancestry. Furthermore, similar associations for genomic ancestry, and individual and parental socioeconomic outcomes were found in a large meta-analysis (Kirkegaard et al., 2017).

Associations of traits with ancestry does not strictly imply genetic causation, but might reflect mere covariance accounted for by another variable. In particular, an often proposed model of racial social inequality is that of racial discrimination, chiefly focused on skin tone (known as *colorism*; (Hunter, 2007; Telles, 2014)). A theoretical path model is shown in Figure 19.

Figure 19: Theoretical path model for simple discrimination model.

In this model, genomic ancestry is statistically associated with both polygenic scores for cognitive ability and variants that code for racially-relevant visual appearance (skin tone, nose shape, hair texture, eye color etc.). People then self-identify as one or more race/ethnicities based on their knowledge of their own ancestry (not necessarily based on genetic data) as well as their own appearance. Their racial appearance and self-identification then lead to racist discrimination from others which finally results in poorer social outcomes, e.g. loss of job opportunity leading to lower incomes, worse living conditions etc. On the other hand, the polygenic scores result in phenotypic cognitive ability which also has an influence on one's eventual socioeconomic outcomes.

The question is then: what is the relative importance of the two pathways to worse socioeconomic outcomes (lower S)? Since most proponents of colorism deny any ancestral differences in polygenic scores for cognitive ability (and other socially valuable traits), they essentially have to embrace some kind of ancestry effect that is not mediated by these, which leaves them with various discrimination models. The simple discrimination model is directly testable if one has measures of the discrimination pathway: skin tone, SIRE, discrimination. Unfortunately, the current dataset does not have a measure of skin tone²⁰, only of SIRE, which were extensively used in this study and generally found not to mediate the ancestry ~ outcome relationship. This leaves us with the question of whether racial appearance is important or perhaps whether other-identified race/ethnicity (OIRE) is.

A crucial piece of data here is whether ancestry and the related SIRE groups actually experience sufficiently different levels of discrimination as would be necessary. Boutwell et al (2017) studied a very large, nationally representative sample of Americans (Add Health) and found that self-reported

20 Due to the high heritability and simplicity of the genetic architecture of skin tone, it can be fairly well predicted from microarray data (Walsh et al., 2017). Unfortunately, we did not have access to the raw genomic data for the present dataset and thus were unable to pursue this approach. Furthermore, it is possible to examine the question solely using genomic data. If the discrimination model holds, then the genetic variants for the appearance traits will be correlated with the genetic variants for cognitive ability and socioeconomic outcomes when global genomic ancestry is held constant. This will not be the case given a pure human capital model assuming no pleiotropy for genetic variants for appearance and human capital traits (Kirkegaard, 2016c).

racial and overall discrimination did not vary as strongly by SIRE as might be expected. Their results are reproduced in Table 20.

SIRE group	Never	Rarely	Sometimes	Often	Sometimes + Often
White (n = 9707)	30	46	19	4	24
Black (n = 2221)	30	38	26	5	32
Hispanic (n = 1578)	33	40	22	5	27
American Indian (n = 126)	25	48	22	5	27
Asian (n = 445)	34	47	17	2	19
Mixed race (n = 561)	26	47	23	4	27
Total (n = 14,638)	30	44	21	4	25

Table 20: Self-reported discrimination experiences (Boutwell et al., 2017, Table 2).

Relative to Whites, Blacks and Hispanics do report more overall self-perceived discrimination. However, since this is self-perceived, it raises the question of whether perhaps less bright people are on average more likely to report more discrimination than they actually experience, or that blacks and hispanics are more likely to interpret poor behavior towards them as being caused by racism when it was in fact due to their own behavior. When we re-analyzed the public part of the same dataset used by Boutwell et al, we found that racial gaps in self-reported discrimination were undetectable ($p > .05$) once IQ was entered as a predictor alongside SIRE (results given in supplementary materials). This result raises concerns about the colorism model as given in the path model above.

Focusing on different parts of the simple discrimination model, Kirkegaard (2017) analyzed multiple large-scale datasets to examine whether skin tone, OIRE, SIRE had any predictive validity for important income once cognitive ability was entered into the model. Again, the results were mostly not in line with the proposed discrimination model because OIRE generally had no detectable effect and when it did, the effect was *positive* not negative. SIRE, on the other hand, was usually negative when modeled along with OIRE, which seems to reflect self-identification processes.

These negative tests of the discrimination model suggests that it needs to be modified. One way to accomplish this in order to bring it into consistency with the reported data is to add a causal path from discrimination to cognitive ability. Thus, not only is discrimination proposed to lower S, but also lower the (arguably) primary cause of S. The results reported by Kirkegaard (2017) can thus be explained as controlling for the mediator. This still leaves the generally positive OIRE ~ income associations in a joint regression with cognitive ability to be explained, as well as the generally near-zero skin tone ~ income.

Furthermore, the discrimination model is implausible for other reasons. First, because many other

variables – self-esteem, depression, mood disorders, social phobia – that one would expect to be more affected by discrimination than cognitive ability and socioeconomic outcomes – do not show large gaps between Whites and African Americans, or even show reverse gaps (Dalliard, 2014). Second, the effect size would have to be implausibly large for such discrimination to account for the large gaps in cognitive ability between SIRE groups. Typical reported effect sizes in social psychology are around $r = .20$ (Richard, Bond, & Stokes-Zoota, 2003; see also Jussim, 2012), which was also found for other fields (Bosco, Aguinis, Singh, Field, & Pierce, 2015; Gignac & Szodorai, 2016). The correlation between self-reported discrimination and IQ is If one takes this as completely causal and uses the African American-European American skin tone gap as 2.5 d, this would account for about 0.38 d, or less than half of the observed gap even under extreme assumptions. Third, the IQ gap between African Americans and White Americans show up very early in life and seems to be 15 IQ points already at age 3, long time before cumulative effects of discrimination could have had much time to set in (Lynn, 2015; Malloy, 2013a). Figure 20 shows a preliminary meta-analysis of the literature. Furthermore, the IQ gap has been remarkably stable close to 15 IQ throughout about 150 years (A'Hearn et al., 2009; Fuerst, 2013; Lynn, 2015; P. L. Roth et al., 2001; Shuey, 1966; but for a contrary view, see Dickens & Flynn, 2006)²¹. This is despite large social changes from overt, legal discrimination against African Americans (slavery, Jim Crow laws) to overt, legal discrimination in favor of African Americans (affirmative action; (Detterman, 2000; Perry, 2015)). This suggests that discrimination is unrelated to the IQ gap.

21 The earliest data we know of is given in A'Hearn et al (2009) who reported literacy and numeracy estimates for Blacks and Whites in the 1850, 1870 and 1900 US censuses. Converting these to z scores based on the normal distribution implies average gaps of 0.75, 1.15 and 0.91 d. These estimates are debatable due to various factors including the freeing of the slaves and large-scale White migration at the time. Notably, the 1850 gap is based on free Blacks only, which were likely a selected group.

Figure 20: Timeline of US Black-White IQ gap. Preliminary meta-analysis. Orange line = linear fit. Blue line = LOESS fit.

A sibling control design can provide a strong test of competing models (Jensen, 1980; Kendler et al., 2016; Malloy, 2013b). Siblings differ only slightly in overall genomic ancestry, but can differ substantially in racial appearance including skin tone because these traits are controlled by relatively few locuses. If we assume that genomic ancestry and polygenic score do not result from the same variants (i.e. no pleiotropy but genetic correlation), then we would not expect to find a notable correlation within sibling pairs of racial appearance and cognitive ability/social outcomes, allowing a direct test of the two competing pathways. A number of studies have utilized sibling control designs in this fashion. Table 21 summarizes the cross-sectional and within-sibling effect sizes.

Cross-section effect size	Between sibling effect size	Ratio

Table 21: Effect sizes of racial appearance in sibling control designs.

The mean ratio of the between sibling effect size is only ... of the cross-sectional one, suggesting weak

causality.

Furthermore, it is possible to combine sibling control designs with genomic data. Such a design has been utilized for estimating the heritability of height by using the fact that siblings are only 50% identical by state (IBS) on average; the standard deviation is $\sim 3.6\%$ points (Visscher et al., 2006). The slight variation around 50% makes it possible to infer the heritability of a trait. A sibling pair that happens to be 52% IBS will be slightly more similar than a sibling pair that happens to be 48% IBS. It is possible to use the same method for ancestry because siblings vary slightly in their ancestries (Malloy, 2013b). Because such small differences are not visible to the eye, they control for any kind of environmental effect by design.

More direct genomic designs are possible. The most straightforward one is simply to calculate the polygenic scores for a large sample of persons with diverse ancestry and see how they are related (or not related). Unfortunately, most GWASs are done solely on people of European ancestry (to avoid population stratification confounding) which has the unfortunate consequence that the resulting predictive models are less useful in non-Europeans, in particularly in Africans (Marigorta & Navarro, 2013; Zanetti & Weale, 2016). The reason for this is a combination of two things. First, the genetic variants discovered in GWASs are generally not causal, but are so-called tag variants, that is, variants that are close on the genome to the causal variants such that their presence in the population is statistically linked (known as linkage disequilibrium; LD). However, the degree to which two variants on the genome are in LD depends to some degree on the population studied because many of these result from random patterns arising over time (genetic drift in LD). The effect of this is thus that a variant that tags a causal variant in Europeans may not do so, or not very well, in other populations, thus reducing validity of the European-derived estimates of the true polygenic scores.

Particularly powerful genomic design that does not require flawless polygenic scores for each population is *admixture mapping* (Coram et al., 2015; Shriner, 2013; Shriver et al., 2003; Winkler, Nelson, & Smith, 2010). It is commonly used for diseases in the medical genetics literature, but can also be used for another other trait and has in fact been used to identify other group differences in adaptive traits such as high-altitude (Jeong et al., 2014) and diving adaptations (Ilardo et al., 2018). The idea with the method is fairly simple: though we may not know exactly where the causal variants are, we know roughly where on the genome they are (i.e. in which gene), and they can plausibly be assumed to be mostly the same loci across populations (Zanetti & Weale, 2016). This means that if we take cases and controls (or high/low trait groups) from a given admixed population where we suspect one ancestry (A) has a higher polygenic score for the trait (e.g. higher disease risk), the cases will have an overall higher proportion of that A ancestry than the controls. However, this increase in ancestry will not be randomly distributed along the genome but will especially concentrated in regions around the causal variants. Thus, the test for the genetic model is whether such local ancestry enrichment can be found for variants found in GWASs for cognitive ability. Because of the very strong polygenicity of cognitive ability, this test will likely require a fairly large sample size.

We were able to strongly disconfirm claims made by various researchers that there are no *statistical relationships* between genomic ancestry and cognitive ability controlling for socially identified racial

groups (Colman, 2016; Nisbett, 2005, 2009; Nisbett et al., 2012; Sternberg, 2005). As discussed above, this does not necessary mean that genetic models are correct because genomic ancestry is probably associated with a large number of other non-genetic variables that run in families for environmental reasons, some of which may be causal. As such, validity ascribed to ancestry in studies such as the present could be due to confounding with these non-genetic variables or reflect ancestry-induced social processes (discrimination, favoritism). One of our analyses included parental S as a predictor of children's cognitive ability, and it was found to be a useful predictor. Inclusion of it into the models reduced the validity of ancestry predictors by 36-48% (cf. Section 3.3.2). As discussed above, a reduction in the effect size is expected on both genetic and non-genetic models. This is because in genetic models, controlling for (holding constant) parental S indirect controls for parental cognitive ability, parental polygenic scores for cognitive ability and ultimately children's polygenic scores for cognitive ability, which then means that the association between genomic ancestry and their polygenic scores is weakened. On non-genetic models, parental S is one among multiple non-genetic causes statistically related to genomic ancestry and holding it constant thus reduces the confounding. Still, we found that controlling for this broad aspect of parental environment – which includes the educational attainment and occupation levels of both parents as well as household income – still left most of the validity of genomic ancestry, consistent with genetic causation. *Scholars who propose purely environmental models should come forward and specifically state which causes they think are responsible for the remaining association with genomic ancestry and their relative importance so they may be properly tested.*

With regards to the effect sizes, the estimated effect size for African and Amerindian ancestry – means of -1.46 and -0.92 across models, i.e. mean IQs of 78 and 86 compared to assumed European of 100 – were in line with Richard Lynn (2015)'s global predictions. Lynn has noted (chapter 4, section 17) that African in Africa score about 70 on a variety of standardized tests, while African Americans score about 85. African Americans enjoy a much higher standard of living than African Africans but also have roughly 20% more European ancestry. Lynn proposed an African ancestry (genotypic) mean IQ of about 80, with the remaining difference of 10 points seen being ascribed to environmental causes, especially nutrition and health. This corresponds to a between group heritability of ~67% between Europeans in European countries and Africans in Africa (~30 points gap, ~20 of which are genetic) and a ~100% between group heritability for European Americans vs. African Americans. Lynn did not provide a best guess for Amerindian genotypic IQ in the analogous section (chapter 11, section 9), however he took 86 as a mean of Amerindian populations of various European admixture and concluded that both environment (chiefly nutrition) and genetics had a major role in the gap. It should be noted, however, that our results were fairly imprecise owing to sample size. If we use the values from the main main for cognitive ability (Table 12), the African genotypic IQ estimate has a 95% confidence interval from 70 to 82 IQ, and for Amerindian from 74 to 92 IQ. Clearly, the values in the present study are too imprecise for exact interpretation of global trends, but they provide a ballpark estimate of the genotypic IQs if one is willing to grant the assumption that the present ancestry results reflect wholly genetic causation.

We conclude that overall the results presented in this paper are more congruent with a simple genetic,

evolutionary model than with a discrimination or shared environment model. However, doubts remain. A single dataset or study cannot and should not settle a debate (Schmidt & Oh, 2016). We encourage researchers to gain access to and use available genomic datasets to do further studies along the lines of the present study (Murray, 2005). As discussed above and despite critics claims (Conley & Fletcher, 2017; Turkheimer, 2017), powerful methods (various sibling designs, admixture mapping, cross-population accurate estimates of polygenic scores) exist that can be used to ascertain whether genetic causes are or are not related to the racial cognitive ability gaps.

Supplementary material and acknowledgments

Supplementary materials (code, figures, detailed numerical output, extra tables etc.):

<https://osf.io/z8dy5/>

References

- A'Hearn, B., Baten, J., & Crayen, D. (2009). Quantifying quantitative literacy: Age heaping and the history of human capital. *The Journal of Economic History*, 69(03), 783–808.
- Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L., ... Jernigan, T. L. (2014). The NIH Toolbox Cognition Battery: Results from a Large Normative Developmental Sample (PING). *Neuropsychology*, 28(1), 1–10.
<https://doi.org/10.1037/neu0000001>
- Andreasen, R. O. (2000). Race: Biological Reality or Social Construct? *Philosophy of Science*, 67, S653–S666. <https://doi.org/10.1086/392853>
- Ash, J., & Gallup, G. G. (2012). Brain size, intelligence, and paleoclimatic variation. In G. Geher & G. Miller (Eds.), *Mating Intelligence: Sex, Relationships, and the Mind's Reproductive System*. Psychology Press.
- Baten, J. (2015, December 14). Numeracy (Total). IISH Dataverse. Retrieved from <https://www.clionfra.eu/Indicators/NumeracyTotal.html>
- Beall, C. M. (2014). Adaptation to High Altitude: Phenotypes and Genotypes. *Annual Review of Anthropology*, 43(1), 251–272. <https://doi.org/10.1146/annurev-anthro-102313-030000>
- Beals, K. L., Smith, C. L., Dodd, S. M., Angel, J. L., Armstrong, E., Blumenberg, B., ... Trinkaus, E. (1984). Brain Size, Cranial Morphology, Climate, and Time Machines [and Comments and Reply]. *Current Anthropology*, 25(3), 301–330. <https://doi.org/10.1086/203138>

- Beauchamp, J. P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences*, 113(28), 7774–7779.
<https://doi.org/10.1073/pnas.1600398113>
- Berg, J. J., Zhang, X., & Coop, G. (2017). Polygenic Adaptation has Impacted Multiple Anthropometric Traits. *BioRxiv*, 167551. <https://doi.org/10.1101/167551>
- Bidulescu, A., Choudhry, S., Musani, S. K., Buxbaum, S. G., Liu, J., Rotimi, C. N., ... Gibbons, G. H. (2014). Associations of adiponectin with individual European ancestry in African Americans: the Jackson Heart Study. *Frontiers in Genetics*, 5. <https://doi.org/10.3389/fgene.2014.00022>
- Borjas, G. J. (2016). *We wanted workers: unraveling the immigration narrative* (First edition). New York: W. W. Norton & Company.
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *The Journal of Applied Psychology*, 100(2), 431–449.
<https://doi.org/10.1037/a0038047>
- Bouchard, T. J. (2013). The Wilson Effect: The Increase in Heritability of IQ With Age. *Twin Research and Human Genetics*, 16(05), 923–930. <https://doi.org/10.1017/thg.2013.54>
- Boutwell, B. B., Nedelec, J. L., Winegard, B., Shackelford, T., Beaver, K. M., Vaughn, M., ... Wright, J. P. (2017). The prevalence of discrimination across racial groups in contemporary America: Results from a nationally representative sample of adults. *PLOS ONE*, 12(8), e0183356.
<https://doi.org/10.1371/journal.pone.0183356>
- Brues, A. M. (1990). *People and Races*. Prospect Heights, Ill.: Waveland Pr Inc.
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*, 96(1), 37–53. <https://doi.org/10.1016/j.ajhg.2014.11.010>
- Byrne, R. (1995). *The Thinking Ape: The Evolutionary Origins of Intelligence*. Oxford ; New York: Oxford University Press.
- Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton university press. Retrieved from <https://www.google.com/books?>

hl=en&lr=&id=FrwNcwKaUKoC&oi=fnd&pg=PP17&dq=Luigi+Luca+Cavalli-Sforza&ots=Hn3VPlGC78&sig=pkJZ8DecXVqSKgfAARtKsFa96Yg

Clark, G. (2007). *A farewell to alms: a brief economic history of the world*. Princeton: Princeton University Press.

Cochran, G., & Harpending, H. (2010). *The 10,000 year explosion: how civilization accelerated human evolution*. New York: Basic Books.

Cofnas, N. (2015). Science Is Not Always “Self-Correcting.” *Foundations of Science*, 21(3), 477–492. <https://doi.org/10.1007/s10699-015-9421-3>

Collared, M., Kemery, M., & Banks, S. (2005). Causes of Toolkit Variation Among Hunter-Gatherers: A Test of Four Competing Hypotheses. *Canadian Journal of Archaeology / Journal Canadien d'Archéologie*, 29(1), 1–19.

Colman, A. M. (2016). Race differences in IQ: Hans Eysenck’s contribution to the debate in the light of subsequent research. *Personality and Individual Differences*, 103, 182–189. <https://doi.org/10.1016/j.paid.2016.04.050>

Conley, D., & Fletcher, J. (2017, June 1). What Both the Left and Right Get Wrong About Race. Retrieved August 31, 2017, from <http://nautil.us/issue/48/chaos/what-both-the-left-and-right-get-wrong-about-race>

Coram, M. A., Candille, S. I., Duan, Q., Chan, K. H. K., Li, Y., Kooperberg, C., ... Tang, H. (2015). Leveraging Multi-ethnic Evidence for Mapping Complex Traits in Minority Populations: An Empirical Bayes Approach. *The American Journal of Human Genetics*, 96(5), 740–752. <https://doi.org/10.1016/j.ajhg.2015.03.008>

Dalliard. (2014). The Elusive X-Factor: A Critique of J. M. Kaplan’s Model of Race and IQ. *Open Differential Psychology*. Retrieved from <https://openpsych.net/paper/27>

DeFries, J. C. (1972). Quantitative Aspects of Genetics and Environment in the Determination of Behavior. In *Genetics, Environment, and Behavior: Implications for Educational Policy* (p. 344). Academic Press.

Detterman, D. K. (2000). Tests, affirmative action in university admissions, and the American way.

- Psychology, Public Policy, and Law*, 6(1), 44–55. <https://doi.org/10.1037/1076-8971.6.1.44>
- Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: evidence from standardization samples. *Psychological Science*, 17(10), 913–920.
<https://doi.org/10.1111/j.1467-9280.2006.01802.x>
- Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences*, 38.
<https://doi.org/10.1017/S0140525X14000430>
- Dunbar, R. I. M. (2003, November 28). The Social Brain: Mind, Language, and Society in Evolutionary Perspective [review-article].
<https://doi.org/10.1146/annurev.anthro.32.061002.093158>
- Dunn, L. M. (1988). Bilingual Hispanic Children on the U.S. Mainland: A Review of Research on Their Cognitive, Linguistic, and Scholastic Development. Emphasizing Studies Involving the English- and Spanish-Language Versions of the Peabody Picture Vocabulary Test--Revised. Retrieved from <http://eric.ed.gov/?id=ED304274>
- Elias, P. M., & Williams, M. L. (2015). Chapter 19 - Evolution of Skin Color. In M. P. Muehlenbein (Ed.), *Basics in Human Evolution* (pp. 273–283). Boston: Academic Press.
<https://doi.org/10.1016/B978-0-12-802652-6.00019-0>
- Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1701), 3801–3808.
- Fan, C. C., Bartsch, H., Schork, A., Chen, C.-H., Wang, Y., Lo, M.-T., ... Dale, A. M. (2015). Modeling the 3D Geometry of the Cortical Surface With Genetic Ancestry. *Current Biology : CB*, 25(15), 1988–1992. <https://doi.org/10.1016/j.cub.2015.06.006>
- Fernandes, H. B. F., Woodley, M. A., & te Nijenhuis, J. (2014). Differences in cognitive abilities among primates are concentrated on G: Phenotypic and phylogenetic comparisons with two meta-analytical databases. *Intelligence*, 46, 311–322. <https://doi.org/10.1016/j.intell.2014.07.007>
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in

stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53(1), 25–44.

<https://doi.org/10.1016/j.jsp.2014.10.002>

Fuerst, J. (2013). Secular Changes in the Black-White Cognitive Ability Gap. Retrieved August 31, 2015, from <http://humanvarieties.org/2013/01/15/secular-changes-in-the-black-white-cognitive-ability-gap/>

Fuerst, J. (2017). Lineage Population: A Concept Needed by an Observer of Nature? *Mankind Quarterly*, 57(4). Retrieved from <http://mankindquarterly.org/archive/issue/57-4/7>

Fuerst, J., & Kirkegaard, E. O. W. (2016a). Admixture in the Americas: Regional and national differences. *Mankind Quarterly*.

Fuerst, J., & Kirkegaard, E. O. W. (2016b). The Genealogy of Differences in the Americas. *Mankind Quarterly*, 56(3), 425–481.

Geary, D. C. (2004). *Origin of Mind: Evolution of Brain, Cognition, and General Intelligence* (1 edition). Washington, DC: Amer Psychological Assn.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Retrieved from <http://www.stat.columbia.edu/~gelman/research/unpublished/>

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., ... Thomas, M. G. (2011). Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1566), 863–877. <https://doi.org/10.1098/rstb.2010.0268>

Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence*, 64, 18–29. <https://doi.org/10.1016/j.intell.2017.06.004>

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>

Giri, A., Edwards, T. L., Hartmann, K. E., Torstenson, E. S., Wellons, M., Schreiner, P. J., & Edwards,

- D. R. V. (2017). African genetic ancestry interacts with body mass index to modify risk for uterine fibroids. *PLOS Genetics*, 13(7), e1006871.
<https://doi.org/10.1371/journal.pgen.1006871>
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79–132.
[https://doi.org/10.1016/S0160-2896\(97\)90014-3](https://doi.org/10.1016/S0160-2896(97)90014-3)
- Gregory, M. D., Kippenhan, J. S., Eisenberg, D. P., Kohn, P. D., Dickinson, D., Mattay, V. S., ... Berman, K. F. (2017). Neanderthal-Derived Genetic Variation Shapes Modern Human Cranium and Brain. *Scientific Reports*, 7(1), 6308. <https://doi.org/10.1038/s41598-017-06587-0>
- Helgason, A., Pálsson, S., Guðbjartsson, D. F., Kristjánsson, Þ., & Stefánsson, K. (2008). An Association Between the Kinship and Fertility of Human Couples. *Science*, 319(5864), 813–816. <https://doi.org/10.1126/science.1150232>
- Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Hill, W. D., Davies, G., McIntosh, A. M., Gale, C. R., & Deary, I. J. (2017). A combined analysis of genetically correlated traits identifies 107 loci associated with intelligence. *BioRxiv*, 160291.
<https://doi.org/10.1101/160291>
- Hunter, M. (2007). The Persistent Problem of Colorism: Skin Tone, Status, and Inequality. *Sociology Compass*, 1(1), 237–254. <https://doi.org/10.1111/j.1751-9020.2007.00006.x>
- Ilardo, M. A., Moltke, I., Korneliussen, T. S., Cheng, J., Stern, A. J., Racimo, F., ... Willerslev, E. (2018). Physiological and Genetic Adaptations to Diving in Sea Nomads. *Cell*, 173(3), 569–580.e15. <https://doi.org/10.1016/j.cell.2018.03.054>
- Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences*, 107(Supplement 2), 8962–8968.
<https://doi.org/10.1073/pnas.0914628107>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: with applications in R*. New York: Springer.
- Jensen, A. R. (1969). Counter Response. *Journal of Social Issues*, 25(4), 219–222.

<https://doi.org/10.1111/j.1540-4560.1969.tb00632.x>

Jensen, A. R. (1973). *Educability and group differences*. New York: Harper & Row.

Jensen, A. R. (1980). Uses of Sibling Data in Educational and Psychological Research. *American Educational Research Journal*, 17(2), 153–170. <https://doi.org/10.2307/1162480>

Jensen, A. R. (1998). *The g factor: the science of mental ability*. Westport, Conn.: Praeger.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., ... Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications*, 5, 3281. <https://doi.org/10.1038/ncomms4281>

Jones, G. (2016). *Hive mind: how your nation's IQ matters so much more than your own*. Stanford, California: Stanford Economics and Finance, an imprint of Stanford University Press.

Joshi, P. K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., ... Wilson, J. F. (2015). Directional dominance on stature and cognition in diverse human populations. *Nature*, 523(7561), 459–462. <https://doi.org/10.1038/nature14618>

Jussim, L. (2012). *Social Perception and Social Reality: Why Accuracy Dominates Bias and Self-Fulfilling Prophecy*. Oxford University Press.

Kan, K.-J., Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2013). On the Nature and Nurture of Intelligence and Specific Cognitive Abilities: The More Heritable, the More Culture Dependent. *Psychological Science*, 24(12), 2420–2428. <https://doi.org/10.1177/0956797613493292>

Kaplan, J. M., & Winther, R. G. (2013). Prisoners of Abstraction? The Theory and Measure of Genetic Variation, and the Very Concept of “Race.” *Biological Theory*, 7(4), 401–412. <https://doi.org/10.1007/s13752-012-0048-0>

Kendler, K. S., Ohlsson, H., Edwards, A. C., Lichtenstein, P., Sundquist, K., & Sundquist, J. (2016). A novel sibling-based design to quantify genetic and shared environmental effects: application to drug abuse, alcohol use disorder and criminal behavior. *Psychological Medicine*, 46(8), 1639–1650. <https://doi.org/10.1017/S003329171500224X>

Kirkegaard, E. O. W. (2014). The international general socioeconomic factor: Factor analyzing

international rankings. *Open Differential Psychology*. Retrieved from <http://openpsych.net/ODP/2014/09/the-international-general-socioeconomic-factor-factor-analyzing-international-rankings/>

Kirkegaard, E. O. W. (2016a). Does sub-European genomic ancestry predict outcomes for US states? *Open Quantitative Sociology & Political Science*. Retrieved from <http://openpsych.net/OQSPS/2016/09/does-sub-european-genomic-ancestry-predict-outcomes-for-us-states/>

Kirkegaard, E. O. W. (2016b). Inequality across US counties: an S factor analysis. *Open Quantitative Sociology & Political Science*. Retrieved from <http://openpsych.net/OQSPS/2016/05/inequality-across-us-counties-an-s-factor-analysis/>

Kirkegaard, E. O. W. (2016c, October 10). Testing colorism using genomic data without phenotypic data. Retrieved February 7, 2017, from <http://emilkirkegaard.dk/en/?p=5974>

Kirkegaard, E. O. W., & Bjerrekær, J. D. (2016). Country of origin and use of social benefits: A large, preregistered study of stereotype accuracy in Denmark. *Open Differential Psychology*. Retrieved from <https://openpsych.net/paper/49>

Kirkegaard, E. O. W., & Fuerst, J. (2017). Admixture in Argentina. *Mankind Quarterly*, 57(4). Retrieved from <http://mankindquarterly.org/archive/issue/57-4/4>

Kirkegaard, E. O. W., Wang, M., & Fuerst, J. (2017). Biogeographic Ancestry and Socioeconomic Outcomes in the Americas: A Meta-Analysis. *Mankind Quarterly*, 57(3). Retrieved from <http://mankindquarterly.org/archive/issue/57-3/10>

Kitcher, P. (2007). Does 'Race' Have a Future? *Philosophy & Public Affairs*, 35(4), 293–317. <https://doi.org/10.1111/j.1088-4963.2007.00115.x>

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., ... Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, 201612113. <https://doi.org/10.1073/pnas.1612113114>

León, F. R., & Burga-León, A. (2018). UV Radiation Associates With State Income Through Complex

- Cognitive Ability in the USA. *Journal of Individual Differences*. Retrieved from <https://econtent.hogrefe.com/doi/abs/10.1027/1614-0001/a000245?journalCode=jid>
- Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., ... Myers, R. M. (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science*, 319(5866), 1100–1104. <https://doi.org/10.1126/science.1153717>
- Lick, D. J., Alter, A. L., & Freeman, J. B. (2017). Superior Pattern Detectors Efficiently Learn, Activate, Apply, and Update Social Stereotypes. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000349>
- Liebler, C. A., & Halpern-Manners, A. (2008). A practical approach to using multiple-race response data: a bridging method for public-use microdata. *Demography*, 45(1), 143–155.
- Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence*. San Francisco: Freeman.
- Lynn, R. (2002). Skin color and intelligence in African Americans. *Population and Environment*, 23(4), 365–375.
- Lynn, R. (2015). *Race Differences in Intelligence* (Revised Edition). Washington Summit Publishers.
- Lynn, R., & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence*, 38(4), 353–360. <https://doi.org/10.1016/j.intell.2010.04.007>
- Lynn, R., & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences* (1st ed.). Ulster Institute for Social Research.
- Lynn, R., & Yadav, P. (2015). Differences in cognitive ability, per capita income, infant mortality, fertility and latitude across the states of India. *Intelligence*, 49, 179–185. <https://doi.org/10.1016/j.intell.2015.01.009>
- Malloy, J. (2013a). The Onset and Development of B-W Ability Differences: Early Infancy to Age 3 (Part 1). Retrieved August 31, 2015, from <http://humanvarieties.org/2013/05/26/the-onset-and-development-of-b-w-ability-differences-early-infancy-to-age-3-part-1/>
- Malloy, J. (2013b, March 30). Cryptic Admixture, Mixed-Race Siblings, & Social Outcomes. Retrieved December 7, 2016, from <http://humanvarieties.org/2013/03/29/cryptic-admixture-mixed-race->

siblings-social-outcomes/

- Manrique Millones, D., Flores-Mendoza, C., & Millones Rivalles, R. (2015). Intelligence in Peru: Students' results in Raven and its relationship to SES. *Intelligence*, 51, 71–78.
<https://doi.org/10.1016/j.intell.2015.05.004>
- Marigorta, U. M., & Navarro, A. (2013). High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. *PLoS Genetics*, 9(6). <https://doi.org/10.1371/journal.pgen.1003566>
- Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., ... Deary, I. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, 44, 26–32. <https://doi.org/10.1016/j.intell.2014.02.006>
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., ... Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528(7583), 499–503. <https://doi.org/10.1038/nature16152>
- Meigs, J. B., Grant, R. W., Piccolo, R., López, L., Florez, J. C., Porneala, B., ... McKinlay, J. B. (2014). Association of African genetic ancestry with fasting glucose and HbA1c levels in non-diabetic individuals: the Boston Area Community Health (BACH) Prediabetes Study. *Diabetologia*, 57(9), 1850–1858. <https://doi.org/10.1007/s00125-014-3301-1>
- Moore, E. G. (1986). Family socialization and the IQ test performance of traditionally and transracially adopted Black children. *Developmental Psychology*, 22(3), 317–326.
<https://doi.org/10.1037/0012-1649.22.3.317>
- Murray, C. (2005). The Inequality Taboo. *Commentary Magazine*. Retrieved from <https://www.commentarymagazine.com/articles/the-inequality-taboo/>
- Natural Resources Conservation Service. (2003, September 8). Major Biomes Map | NRCS Soils. Retrieved August 29, 2017, from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054002
- Nisbett, R. E. (2005). Heredity, environment, and race differences in IQ: A commentary on Rushton and Jensen (2005). *Psychology, Public Policy, and Law*, 11(2), 302–310.
<https://doi.org/10.1037/1076-8971.11.2.302>

- Nisbett, R. E. (2009). *Intelligence and how to get it: why schools and cultures count*. New York: W.W. Norton & Co.
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Group differences in IQ are best understood as environmental in origin. *American Psychologist*, 67(6), 503–504. <https://doi.org/10.1037/a0029772>
- Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., ... Sowell, E. R. (2015). Family income, parental education and brain structure in children and adolescents. *Nature Neuroscience*, 18(5), 773–778. <https://doi.org/10.1038/nn.3983>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour*, 78(4), 1001–1010. <https://doi.org/10.1016/j.anbehav.2009.06.033>
- Perry, M. J. (2015, January 4). Acceptance rates at US medical schools in 2014 reveal ongoing discrimination against Asian-Americans and whites. Retrieved December 21, 2015, from <https://www.aei.org/publication/acceptance-rates-us-medical-schools-2014-reveal-ongoing-racial-profiling-affirmative-discrimination-blacks-hispanics/>
- Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence*, 53, 43–50. <https://doi.org/10.1016/j.intell.2015.08.008>
- Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences*, 99(7), 4436–4441. <https://doi.org/10.1073/pnas.062041299>
- Reich, D. (2018). *Who We Are and How We Got Here: ancient dna and the new science of the human past*. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1737810>
- Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One Hundred Years of Social Psychology Quantitatively Described. *Review of General Psychology*, 7(4), 331–363.

<https://doi.org/10.1037/1089-2680.7.4.331>

Richwine, J. (2009). *IQ and immigration policy* (Ph.D.). Harvard University Cambridge, MA.

Retrieved from <https://isgp-studies.com/miscellaneous/race-and-iq/2009-jason-richwine-harvard-dissertation-white-asian-black-hispanic-india-iq.pdf>

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., ... Koellinger, P. D. (2014).

Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences*, 111(38), 13790–13794.

<https://doi.org/10.1073/pnas.1404623111>

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: the homogeneity

of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21(5), 667–706. <https://doi.org/10.1002/per.634>

Rindermann, H. (2018). *Cognitive capitalism: human capital and the wellbeing of nations*. Cambridge,

United Kingdom ; New York, NY: University Printing House.

Rindermann, H., Becker, D., & Coyle, T. (2016). Survey of Expert Opinion on Intelligence: Causes of

International Differences in Cognitive Ability Tests. *Educational Psychology*, 7, 399.

<https://doi.org/10.3389/fpsyg.2016.00399>

Rindermann, H., & Thompson, J. (2014). The cognitive competences of immigrant and native students

across the world: an analysis of gaps, possible causes and impact. *Journal of Biosocial Science*, 1–28. <https://doi.org/10.1017/S0021932014000480>

Robinson, M. R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T., Shakhbazov, K., ...

Visscher, P. M. (2015). Population genetic differentiation of height and body mass index across Europe. *Nature Genetics*, 47(11), 1357–1362. <https://doi.org/10.1038/ng.3401>

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic Group Differences in

Cognitive Ability in Employment and Educational Settings: A Meta-Analysis. *Personnel Psychology*, 54(2), 297–330. <https://doi.org/10.1111/j.1744-6570.2001.tb00094.x>

Roth, T. C., & Pravosudov, V. V. (2009). Hippocampal volumes and neuron numbers increase along a

gradient of environmental harshness: a large-scale comparison. *Proceedings of the Royal*

- Society B: Biological Sciences*, 276(1656), 401–405. <https://doi.org/10.1098/rspb.2008.1184>
- Rowe, D. C., & Rodgers, J. L. (1997). Poverty and Behavior: Are Environmental Measures Nature and Nurture? *Developmental Review*, 17(3), 358–375. <https://doi.org/10.1006/drev.1997.0434>
- Rushton, J. P. (2000). *Race, evolution, and behavior: a life history perspective*. Port Huron, MI: Charles Darwin Research Institute.
- Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11(2), 235–294. <https://doi.org/10.1037/1076-8971.11.2.235>
- Sandefur, J. (2016). *Internationally Comparable Mathematics Scores for Fourteen African Countries* (SSRN Scholarly Paper No. ID 2893768). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=2893768>
- Sayol, F., Maspons, J., Lapiedra, O., Iwaniuk, A. N., Székely, T., & Sol, D. (2016). Environmental variation and the evolution of large brains in birds. *Nature Communications*, 7, 13971. <https://doi.org/10.1038/ncomms13971>
- Schaik, C. P. van, & Burkart, J. M. (2011). Social learning and evolution: the cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1567), 1008–1016. <https://doi.org/10.1098/rstb.2010.0304>
- Schmidt, F. L., & Oh, I.-S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, 4(1), 32–37. <https://doi.org/10.1037/arc0000029>
- Schoenemann, P. T. (2006). Evolution of the Size and Functional Areas of the Human Brain. *Annual Review of Anthropology*, 35(1), 379–406. <https://doi.org/10.1146/annurev.anthro.35.081705.123210>
- Schrider, D. R., & Kern, A. D. (2017). Soft Sweeps Are the Dominant Mode of Adaptation in the Human Genome. *Molecular Biology and Evolution*, 34(8), 1863–1877. <https://doi.org/10.1093/molbev/msx154>
- Sesardić, N. (2005). *Making sense of heritability*. Cambridge, UK; New York: Cambridge University

- Press. Retrieved from <http://public.eblib.com/choice/publicfullrecord.aspx?p=241083>
- Sesardic, N. (2010). Race: a social destruction of a biological concept. *Biology & Philosophy*, 25(2), 143–162. <https://doi.org/10.1007/s10539-009-9193-7>
- Shriner, D. (2013). Overview of Admixture Mapping. *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.], CHAPTER, Unit1.23*. <https://doi.org/10.1002/0471142905.hg0123s76>
- Shriver, M. D., Parra, E. J., Dios, S., Bonilla, C., Norton, H., Jovel, C., ... Kittles, R. A. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*, 112(4), 387–399. <https://doi.org/10.1007/s00439-002-0896-y>
- Shuey, A. M. (1966). *The Testing of Negro Intelligence* (2nd ed.). Social Science Press.
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., & Nosek, B. A. (In review). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. Retrieved from <https://osf.io/qix4g/>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simonti, C. N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D. S., Chisholm, R. L., ... Capra, J. A. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science*, 351(6274), 737–741. <https://doi.org/10.1126/science.aad2149>
- Sol, D., Garcia, N., Iwaniuk, A., Davis, K., Meade, A., Boyle, W. A., & Székely, T. (2010). Evolutionary Divergence in Brain Size between Migratory and Resident Birds. *PLOS ONE*, 5(3), e9617. <https://doi.org/10.1371/journal.pone.0009617>
- Sol, D., Lefebvre, L., & Rodríguez-Teijeiro, J. D. (2005). Brain size, innovative propensity and migratory behaviour in temperate Palaearctic birds. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1571), 1433–1441. <https://doi.org/10.1098/rspb.2005.3099>
- Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1611), 763–769.

<https://doi.org/10.1098/rspb.2006.3765>

Sowell, T. (1995). *Race and culture: a world view*. New York: Basic Books.

Sternberg, R. J. (2005). There are no public-policy implicatons: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law*, 11(2), 295–301. <https://doi.org/10.1037/1076-8971.11.2.295>

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35(5), 401–426. <https://doi.org/10.1016/j.intell.2006.09.004>

Suzuki, L., & Aronson, J. (2005). The cultural malleability of intelligence and its impact on the racial/ethnic hierarchy. *Psychology, Public Policy, and Law*, 11(2), 320–327. <https://doi.org/10.1037/1076-8971.11.2.320>

Telles, E. E. (2014). *Pigmentocracies: ethnicity, race, and color in Latin America* (1 Edition). Chapel Hill, NC: University of North Carolina Press.

Templ, M., Alfons, A., Kowarik, A., & Prantner, B. (2015, February 19). VIM: Visualization and Imputation of Missing Values. CRAN. Retrieved from <http://cran.r-project.org/web/packages/VIM/index.html>

Templ, M., Kowarik, A., & Filzmoser, P. (2011). Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, 55(10), 2793–2806. <https://doi.org/10.1016/j.csda.2011.04.012>

Templer, D. I., & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. *Intelligence*, 34(2), 121–139. <https://doi.org/10.1016/j.intell.2005.04.002>

Templer, D. I., & Stephens, J. S. (2014). The relationship between IQ and climatic variables in African and Eurasian countries. *Intelligence*, 46, 169–178. <https://doi.org/10.1016/j.intell.2014.06.001>

Tizard, B., Cooperman, O., Joseph, A., & Tizard, J. (1972). Environmental Effects on Language Development: A Study of Young Children in Long-Stay Residential Nurseries. *Child Development*, 43(2), 337–358. <https://doi.org/10.2307/1127540>

Turchin, M. C., Chiang, C. W., Palmer, C. D., Sankararaman, S., Reich, D., Genetic Investigation of

- ANthropometric Traits (GIANT) Consortium, & Hirschhorn, J. N. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics*, 44(9), 1015–1019. <https://doi.org/10.1038/ng.2368>
- Turkheimer, E. (2007). Race and IQ. *Cato Unbound*. Retrieved from <http://www.cato-unbound.org/2007/11/21/eric-turkheimer/race-iq>
- Turkheimer, E. (2017, June 2). Origin of Race Differences in Intelligence is Not a Scientific Question. Retrieved August 31, 2017, from <http://www.geneticshumanagency.org/gha/origin-of-race-differences-in-intelligence-is-not-a-scientific-question/>
- Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., ... Martin, N. G. (2006). Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLOS Genetics*, 2(3), e41. <https://doi.org/10.1371/journal.pgen.0020041>
- Walsh, S., Chaitanya, L., Breslin, K., Muralidharan, C., Bronikowska, A., Pospiech, E., ... Kayser, M. (2017). Global skin colour prediction from DNA. *Human Genetics*, 136(7), 847–863. <https://doi.org/10.1007/s00439-017-1808-5>
- Westfall, J., & Yarkoni, T. (2016). Statistically Controlling for Confounding Constructs Is Harder than You Think. *PLOS ONE*, 11(3), e0152719. <https://doi.org/10.1371/journal.pone.0152719>
- Wicherts, J. M., Veldkamp, C. L. S., Augusteyn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01832>
- Winkler, C. A., Nelson, G. W., & Smith, M. W. (2010). Admixture mapping comes of age. *Annual Review of Genomics and Human Genetics*, 11, 65–89. <https://doi.org/10.1146/annurev-genom-082509-141523>
- Woodley, M. A. (2010). Is Homo sapiens polytypic? Human taxonomic diversity and its implications. *Medical Hypotheses*, 74(1), 195–201. <https://doi.org/10.1016/j.mehy.2009.07.046>
- Woodley, M. A., Younuskunju, S., Balan, B., & Piffer, D. (2017). Holocene Selection for Variants

Associated With General Cognitive Ability: Comparing Ancient and Modern Genomes. *Twin Research and Human Genetics*, 20(4), 271–280. <https://doi.org/10.1017/thg.2017.37>

Zabaneh, D., Krapohl, E., Gaspar, H. A., Curtis, C., Lee, S. H., Patel, H., ... Breen, G. (2017). A genome-wide association study for extremely high intelligence. *Molecular Psychiatry*. <https://doi.org/10.1038/mp.2017.121>

Zanetti, D., & Weale, M. E. (2016). True causal effect size heterogeneity is not required to explain trans-ethnic differences in GWAS signals. *BioRxiv*, 085092. <https://doi.org/10.1101/085092>

Zou, J. Y., Park, D. S., Burchard, E. G., Torgerson, D. G., Pino-Yanes, M., Song, Y. S., ... Zaitlen, N. (2015). Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. *Proceedings of the National Academy of Sciences*, 112(44), 13621–13626. <https://doi.org/10.1073/pnas.1501741112>